DOI:
10.1371/journal.pone.0177977

Document Version
Publisher's PDF, also known as Version of record

PLOS ONE

RESEARCH ARTICLE

A dual-strategy expression screen for candidate connectivity labels in the developing thalamus

Olivia Bibollet-Bahena[1], Tatsuya Okafuji[1], Karsten Hokamp[1], Guy Tear[2], Kevin J. Mitchell[1,3]*

1 Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, **2** Department of Developmental Neurobiology, New Hunt's House, Guy's Campus, King's College, London, United Kingdom, **3** Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

* Kevin.Mitchell@tcd.ie

Citation: Bibollet-Bahena O, Okafuji T, Hokamp K, Tear G, Mitchell KJ (2017) A dual-strategy expression screen for candidate connectivity labels in the developing thalamus. PLoS ONE 12(5): e0177977. https://doi.org/10.1371/journal.pone.0177977

Editor: Izumi Sugihara, Tokyo Medical and Dental University, JAPAN

Received: October 13, 2016

Accepted: May 5, 2017

Published: May 30, 2017

Data Availability Statement: All data are contained within the paper and Supporting Information.

Funding: This work was supported by grants to KJM and GT from the Wellcome Trust (075264/A/04/Z) and Science Foundation Ireland (07/IN.1/B969 and 09/IN.1/B2614) and to OBB from the Fonds de recherche Santé Québec (23980).

Competing interests: The authors have declared that no competing interests exist.

Abstract

The thalamus or "inner chamber" of the brain is divided into ~30 discrete nuclei, with highly specific patterns of afferent and efferent connectivity. To identify genes that may direct these patterns of connectivity, we used two strategies. First, we used a bioinformatics pipeline to survey the predicted proteomes of nematode, fruitfly, mouse and human for extracellular proteins containing any of a list of motifs found in known guidance or connectivity molecules. Second, we performed clustering analyses on the Allen Developing Mouse Brain Atlas data to identify genes encoding surface proteins expressed with temporal profiles similar to known guidance or connectivity molecules. In both cases, we then screened the resultant genes for selective expression patterns in the developing thalamus. These approaches identified 82 candidate connectivity labels in the developing thalamus. These molecules include many members of the Ephrin, Eph-receptor, cadherin, protocadherin, semaphorin, plexin, Odz/teneurin, Neto, cerebellin, calsyntenin and Netrin-G families, as well as diverse members of the immunoglobulin (Ig) and leucine-rich receptor (LRR) superfamilies, receptor tyrosine kinases and phosphatases, a variety of growth factors and receptors, and a large number of miscellaneous membrane-associated or secreted proteins not previously implicated in axonal guidance or neuronal connectivity. The diversity of their expression patterns indicates that thalamic nuclei are highly differentiated from each other, with each one displaying a unique repertoire of these molecules, consistent with a combinatorial logic to the specification of thalamic connectivity.

Introduction

The thalamus, or "inner chamber" of the brain, is a crucial nexus in the brain's circuitry. It is not only a relay station that conveys sensory information from the periphery to the cerebral cortex, it is also a conduit for cortico-cortical communication [1], as well as a central node in pathways controlling action selection, through cortico-baso-thalamo-cortical loops [2]. In

addition, the thalamus is interconnected with many other brain structures, including hippocampus, hypothalamus, amygdala, inferior and superior colliculus [3], cerebellar nuclei [4], substantia nigra, brainstem, spinal cord and many others. The thalamus proper, or dorsal thalamus, is also intimately interconnected with the prethalamic reticular nucleus [5], which provides inhibitory regulation of information flow through thalamus.

A striking characteristic of the thalamus is its subdivision into ~30 discrete nuclei, which subserve distinct functions and which have highly selective connectivity patterns with the structures mentioned above, most obviously with specific cortical areas. Some nuclei, such as those conveying primary sensory information, project quite selectively to one or a small number of cortical areas, while others, which integrate signals from multiple sources, project in a more diffuse manner across larger areas of cortex. Within each nucleus there are also varying proportions of distinct cell types that either project to specific areas, driving receptive field properties in input layers of cortex, such as layer 4, or that project more widely across cortex and provide modulatory inputs, for example to layer 1 [6, 7].

Different thalamic nuclei thus act as discrete targets for innervation from a wide array of sources, and, in turn, project their axons to numerous other structures with areal, laminar and cell-type specificity. In order to establish these connections, growing axons must be guided to appropriate regions, must recognise the appropriate target and avoid inappropriate ones, and must make the right kinds of synapses on the right kinds of cells. These processes are mediated in general by surface and secreted proteins, which act as signals and receptors, enabling cellular recognition for pathfinding and target selection.

Numerous protein families have been identified as playing important roles in thalamic axon guidance and connectivity. Many of these, including Ephrins/Eph-receptors, Netrins/DCC/Unc5s, Slits/Robos, Neuregulin-1/Erbb4, secreted semaphorins/ neuropilins and L1 cell adhesion molecules, mediate general processes such as avoidance of the hypothalamus, projection into the internal capsule and topographic organisation of thalamocortical and/or corticothalamic projections through this intermediate target region (reviewed in [8, 9]). By contrast, very few molecules have been found so far that mediate more specific connectivity relationships of particular thalamic nuclei, such as Cdh6 [10], or that control sub-organisation of projections within thalamic nuclei, such as Lrp8 [11] or Ten-m3/Odz3 [12], which also regulates topography of projections to striatum [13].

At earlier stages, a number of studies have described the combinatorial expression patterns of patterning molecules or transcription factors across the thalamus, many of which are expressed selectively in some nuclei and not others [14–18]. Some have also documented the differential expression of selected surface proteins across the thalamus [15, 19, 20], and in some cases direct links have been shown between transcription factor and surface molecule expression [21, 22]. However, these analyses have not been performed in a systematic or comprehensive fashion. Thus, while we have learned a lot about how the developing thalamus is patterned and how the fates of different nuclei are specified, we still know relatively little about the combinatorial code of surface molecules that specifies nuclear connectivity.

Here, we describe two parallel approaches to identify genes encoding surface or secreted proteins expressed in discrete patterns across the developing thalamus. First, we screened the predicted proteomes of human, mouse, fly and worm for conserved genes encoding predicted transmembrane proteins with any of a number of protein motifs commonly found in axon guidance molecules. These genes were screened by *in situ* hybridization to find those expressed in selective or differential patterns across the neonatal thalamus. Second, we analysed the Allen Developing Mouse Brain Atlas (devABA) database for genes encoding surface or secreted proteins, which showed temporal expression profiles similar to known genes for axon guidance or synaptic connectivity. The expression patterns of these genes were then examined to identify

selectively or differentially expressed genes at mid or late embryonic stages. Together, these approaches have identified 82 genes encoding candidate connectivity labels in the developing thalamus. The expression patterns of these genes are highly diverse, such that individual thalamic nuclei express distinct repertoires of these surface molecules, consistent with a combinatorial logic to the specification of connectivity.

Results

A bioinformatics and expression screen to identify conserved candidate connectivity labels

The proteomes of mammals contain many predicted proteins of unknown function. We were interested to discover genes encoding predicted transmembrane proteins that contain any of a number of protein motifs found in known axon guidance molecules. As many known axon guidance molecules are highly conserved from vertebrates to invertebrates, we further concentrated on proteins that had predicted orthologues in the fruitfly *Drosophila melanogaster* and/or in the nematode *Caenorhabditis elegans*.

To identify such proteins, we used protein localisation and motif searches to annotate the predicted proteomes of worm, fly, mouse and human. We also generated a matrix of pairwise BLAST scores across all the members of these proteomes and clustered them using the TRIBE-MCL algorithm, as previously described [23]. We screened through the resultant outputs to find clusters with mammalian and fly or worm members encoding predicted transmembrane proteins, which also contained any of the following motifs: immunoglobulin (Ig) domain, fibronectin type 3 domain, cadherin domain, leucine-rich repeat (LRR), EGF repeat, CUB domain, sema domain, or plexin repeat. We were particularly interested in discovering novel axon guidance or connectivity cues and so excluded genes or gene families where such a function had already been demonstrated in the fly or mouse at the time the screen was performed (in 2006). These already known gene families included, among others: *DCC* and *Neogenin*, Robos, Slits, L1CAMs, Contactins, Eph-receptors, Ephrins, FGFs and FGFRs, receptor protein tyrosine phosphatases, Cadherins, Protocadherins, Semaphorins, Plexins and Integrins. (Leucine-rich repeat family members have been previously described [23]). Using this approach, we prioritized a set of 42 less well-investigated genes for expression screening (S1 Table).

We performed *in situ* hybridization for these genes on P0 mouse brain sections, as previously described for genes encoding LRR proteins [23]. Here we concentrate on members of other gene families that show selective or differential expression across the thalamus (Table 1 Column A).

Immunoglobulin superfamily members. We found three Ig superfamily members with selective expression across the thalamus: *Igsf9b*, *Kirrel3*, and *Sdk2* (Fig 1; S1 Fig). *Igsf9b* (Immunoglobulin Superfamily, Member 9B) is an orthologue of the *Drosophila* gene *turtle*, which has been implicated in various aspects of axon guidance (reviewed in [24]). Both *Igsf9b* and *Igsf9* have recently been shown to play roles in inhibitory synapse development in mammals [25, 26]. *Igsf9b* is more strongly expressed in some thalamic nuclei than others, but we detected only weak expression of *Igsf9* in thalamus (not shown).

Kirrel3 (Kin Of IRRE Like 3 (*Drosophila*), also known as *NEPH2*) is an orthologue of the fly gene *irregular chiasm* (*IrreC*), which functions in axonal pathfinding and target selection (reviewed in [27]) and the *C. elegans* gene *SYG-1*, which is implicated in selective synaptogenesis [28]. *Kirrel3* has recently been implicated in axon guidance in the vomeronasal system [29] and synapse formation in the hippocampus [30], but functions in the developing thalamus

Table 1. Candidate connectivity labels.

A. Conserved candidate connectivity labels	B. devABA candidate connectivity labels		
Clstn1	Alcam	EfnA2	Mdga1
Clstn2	Astn2	EfnA5	Nrn
Igsf9b	Bmp3	EfnB3	Ntng1
Kirrel3	Cadm1	EphA1	Ntng2
Neto1	Cbln2	EphA3	Ntrk2
Neto2	Cbln4	EphA4	Ntrk3
Odz1	Cd47	EphA6	Pcdh1
Odz2	Cdh2	EphA8	Pcdh10
Odz3	Cdh4	EphA10	Pcdh11X
Odz4	Cdh6	EphB1	Pcdh19
Sdk2	Cdh7	EphB2	Pcdh21
	Cdh8	EphB6	PlxnA1
	Cdh9	Fat3	PlxnA2
	Cdh10	Flt3	PlxnC1
	Cdh11	Fzd7	Ptpru
	Cdh12	Gfra1	Ret
	Cdh13	Gfra2	Rtn4rl1
	Cdh24	Gpc3	Sema3F
	Clstn1	Igfbp5	Sema6A
	Clstn2	Inhba	Sema7A
	Cntn6	Kit	Odz3
	Cntnap4	Lgi2	Tgfb2
	Dlk1	Lrp8	Trp53i11
	Dner	Lrrn3	Vgf
		Lypd1	Wif1

Table 1 lists all genes for which we present expression patterns. Clstn1, Clstn2 and Odz3 appear on both lists. The total number of candidate connectivity labels with expression documented here is 82.

https://doi.org/10.1371/journal.pone.0177977.t001

have not been described. *Kirrel1* and *Kirrel2* both showed only weak/background expression in thalamus (not shown).

Sdk2 (*Sidekick2*) is an orthologue of the fly gene *sidekick*, which regulates cellular differentiation in the fly eye [31]. An important role for Sidekicks has also been demonstrated in the vertebrate retina in specifying lamina formation and synaptic connectivity [32, 33]. *Sdk2* is quite selectively expressed across thalamic nuclei while *Sdk1* is expressed at high levels across the whole dorsal thalamus (not shown).

***Odz* (*Teneurin*) genes.** We found that all four members of the Odz/Teneurin family are expressed in selective fashion across the developing thalamus at P0 (Fig 2; S2 Fig) and at E15 (S3 and S4 Figs). These genes are orthologues of the fly *odd oz/teneurin* genes *ten-a* and *ten-m*, which function in synaptic connectivity in the olfactory and neuromuscular systems (reviewed in [34]). In mammals, they have both homophilic and heterophilic interactions. *Odz2* (*Ten-m2*) and *Odz4* (*Ten-m4*) are expressed in similar, but not identical patterns, and *Odz3* (*Ten-m3*) expression overlaps substantially with them. *Odz1* (*Ten-m1*) is expressed in a broadly complementary pattern to the other three genes. These genes also show graded expression across cortex and striatum at E15 and P0 (S2 and S4 Figs). Roles for *Odz2* and *Odz3* have been demonstrated in establishing the topography and segregation of ipsilateral and contralateral

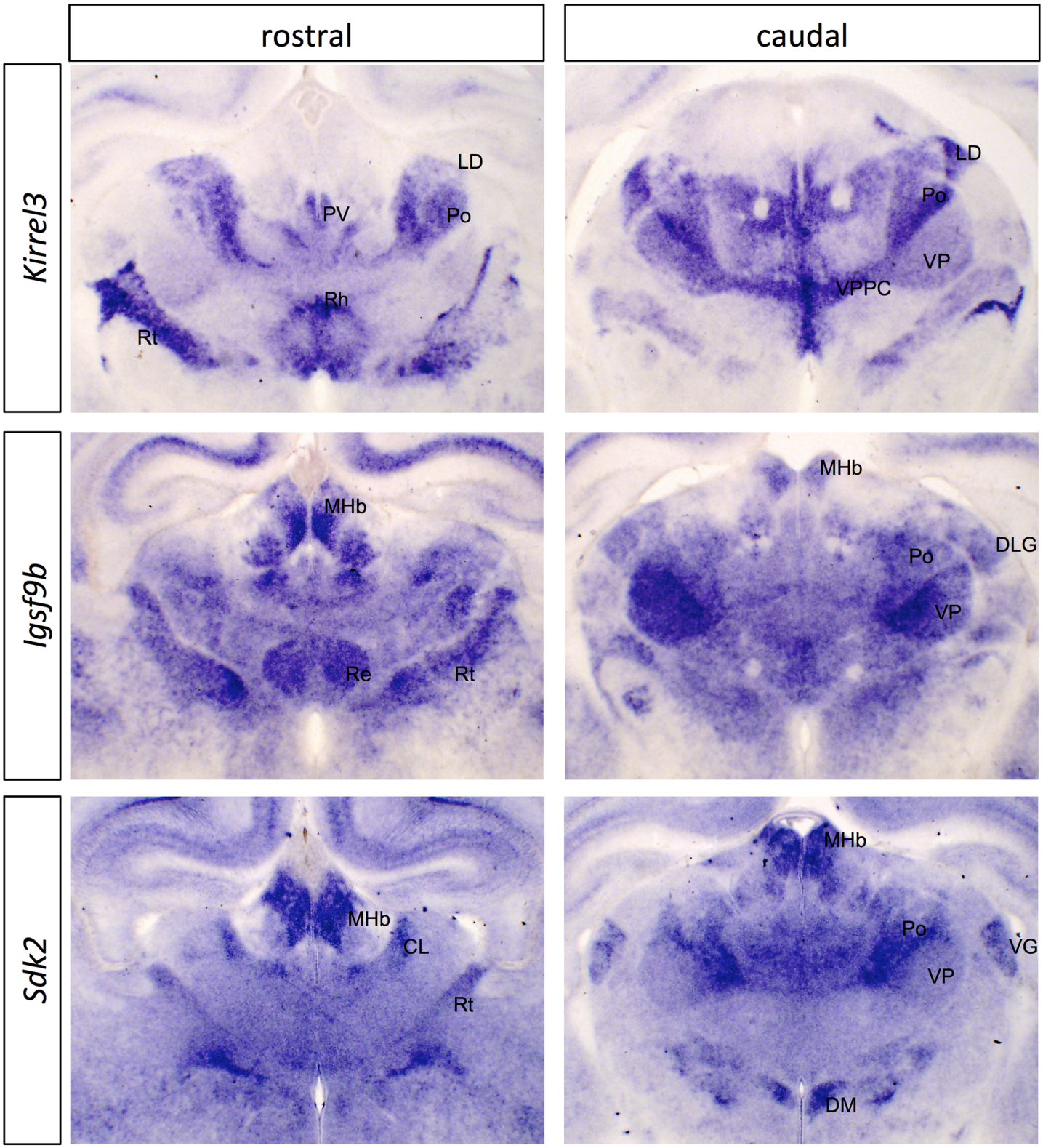

Fig 1. *In situ* hybridization patterns for immunoglobulin superfamily genes in dorsal thalamus at P0. Two coronal sections are shown for *Kirrel3*, *Igsf9b* and *Sdk2*, one rostral and one more caudal. The entire corresponding sections are shown in S1 Fig, for context. CL, centrolateral nucleus; DLG, dorsolateral geniculate nucleus; DM, dorosmedial hypothalamic nucleus; LD, laterodorsal nucleus; MHb, medial habenula; Po, posterior thalamic nuclear group; PV, paraventricular nucleus; Re, reuniens nucleus; Rh, rhomboid nucleus; Rt, reticular nucleus; VG, ventral geniculate nucleus; VP, ventral posterior nucleus; VPPC, ventral posterior nucleus parvicellular part. Scale bar: 500 μm.

https://doi.org/10.1371/journal.pone.0177977.g001

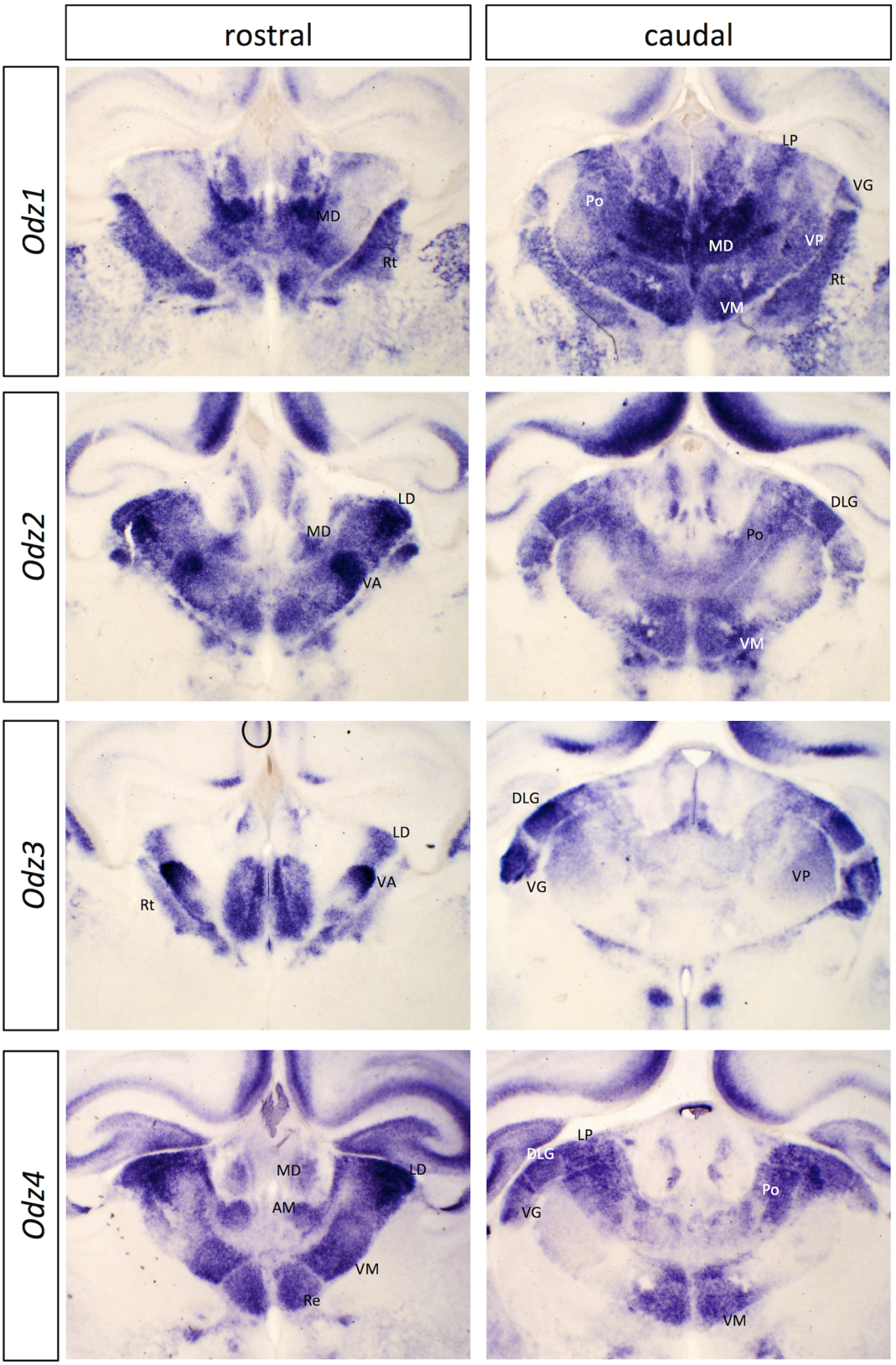

Fig 2. *In situ* hybridization patterns for Odz family genes in dorsal thalamus at P0. Two coronal sections are shown for *Odz1*, *Odz2*, *Odz3* and *Odz4*, one rostral and one more caudal. The entire corresponding sections are

shown in S2 Fig, for context. AM, anteromedial nucleus; DLG, dorsolateral geniculate nucleus; LD, laterodorsal nucleus; MD, mediodorsal nucleus; Po, posterior thalamic nuclear group; Re, reuniens nucleus; Rt, reticular nucleus; VA, ventral anterior nucleus; VG, ventral geniculate nucleus; VM, ventromedial nucleus; VP, ventral posterior nucleus. Scale bar: 500 μm.

https://doi.org/10.1371/journal.pone.0177977.g002

afferents in the dLGN and of thalamic afferents in visual cortex (reviewed in [35]). More recently, *Odz3* has also been shown to control topography of thalamostriatal projections [13].

***Neto* genes.** The neuropilin and tolloid-like genes *Neto1* and *Neto2*, orthologues of the fly gene *neto*, also showed discrete expression in the thalamus, with *Neto2* more widely expressed (Fig 3; S5 Fig). In some parts of the thalamus, their expression was strikingly complementary, particularly in the ventrobasal complex, which expresses high levels of *Neto1* but not *Neto2*, while many other nuclei show the opposite pattern. The Neto proteins function as auxiliary subunits for kainate receptors in mammals (reviewed in [36]) and in glutamate receptor clustering at the neuromuscular junction in flies [37]. No role in axon guidance or synaptic connectivity has been demonstrated for these genes, but their early expression (S6 Fig) and protein similarity to neuropilins would certainly be consistent with such a function.

Calsyntenins. Calsyntenins are transmembrane proteins with two extracellular cadherin domains [38]. They have recently been implicated in synaptogenesis through interactions with neurexins [39]. As with previous reports, we find *Clstn1* to be broadly (but not ubiquitously) expressed across the thalamus [38, 39], while *Clstn2* is expressed more selectively (Fig 4), in contrast to report by Hintsch *et al.* of only background expression. *Clstn3* mRNA was not detectable.

Clustering analyses of developmental gene expression patterns in the thalamus

As a complementary strategy to identify candidate connectivity labels, we analysed data from the devABA, extending our search beyond genes conserved between vertebrates and invertebrates. The devABA provides qualitative and quantitative data on expression of 2002 genes based on mRNA *in situ* hybridization. These genes were chosen based on functional relevance to brain development or disorders of the brain. The expression density values for each gene in the atlas have been mapped to a three-dimensional digital template, with individual voxels attributed to different brain regions. We therefore extracted the expression density values for voxels attributed to the dorsal thalamus for 2002 genes at each of seven ages: embryonic day 11.5 (E11.5), E13.5, E15.5, E18.5, postnatal day 4 (P4), P14 and P28 (S2 Table). These time-points are developmentally relevant to regional specification/patterning (E11.5), axon path-finding (E13.5, E15.5 and E18.5), synaptogenesis (P4 and P14) and cortical plasticity (P28). Because we were more interested in comparing temporal profiles than relative levels across genes, we normalised these data for each gene by dividing by the average expression value for that gene across all ages (S3 Table Columns B-H).

In order to find genes with similar temporal expression profiles we used *k*-means clustering, which clusters elements of a matrix into a user-defined number of clusters (the *k* value). Since there is not necessarily a "correct" number of clusters, we performed *k*-means clustering with values of *k* from 6 to 18 and examined the flux in clustering outputs across these levels. As the input value of *k* increased, clusters tended to get subdivided as opposed to reshuffled, giving a roughly hierarchical organisation (S3 Table Columns I-U; Fig 5). We selected $k = 10$ for further data analysis, since the data segregated into clusters differentially peaking at all embryonic and postnatal time-points, while leaving large enough clusters for statistical analyses of enrichment (Fig 6). Lower *k* values gave poorer separation across the embryonic and early postnatal

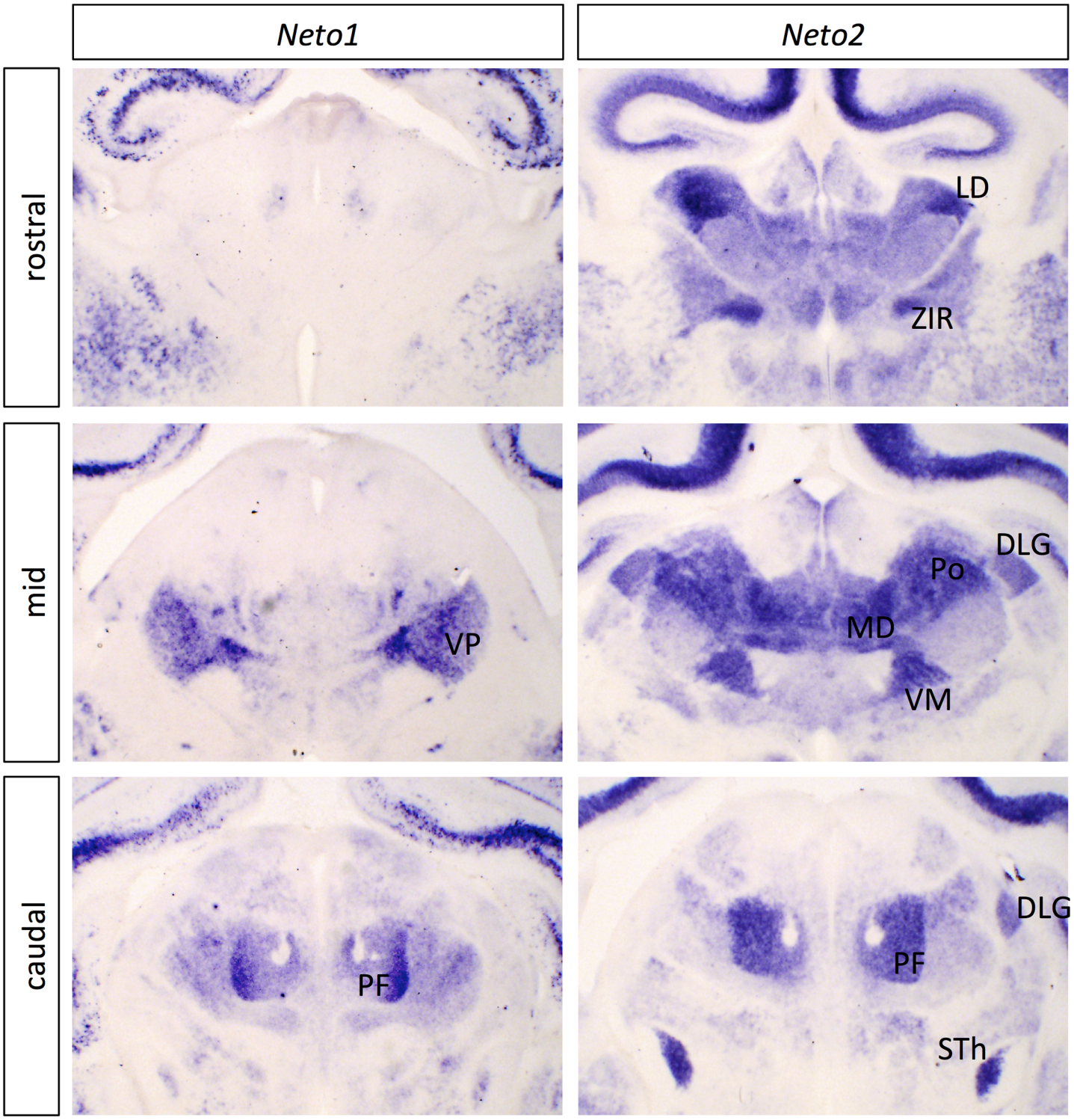

Fig 3. *In situ* hybridization patterns for Neto family genes in dorsal thalamus at P0. Three coronal sections are shown for *Neto1* and *Neto2*, one rostral, one at an intermediate level (mid) and one more caudal. The entire corresponding sections are shown in S5 Fig, for context. Expression of the *Neto* genes is notably complementary in some places (compare mid sections) and overlapping in others (e.g., PF). DLG, dorsolateral geniculate nucleus; LD, laterodorsal nucleus; MD, mediodorsal nucleus; PF, parafscicular nucleus; Po, posterior thalamic nuclear group; STh, subthalamic nucleus; VM, ventromedial nucleus; VP, ventral posterior nucleus; ZIR, zona incerta rostral part. Scale bar: 500 μm.

https://doi.org/10.1371/journal.pone.0177977.g003

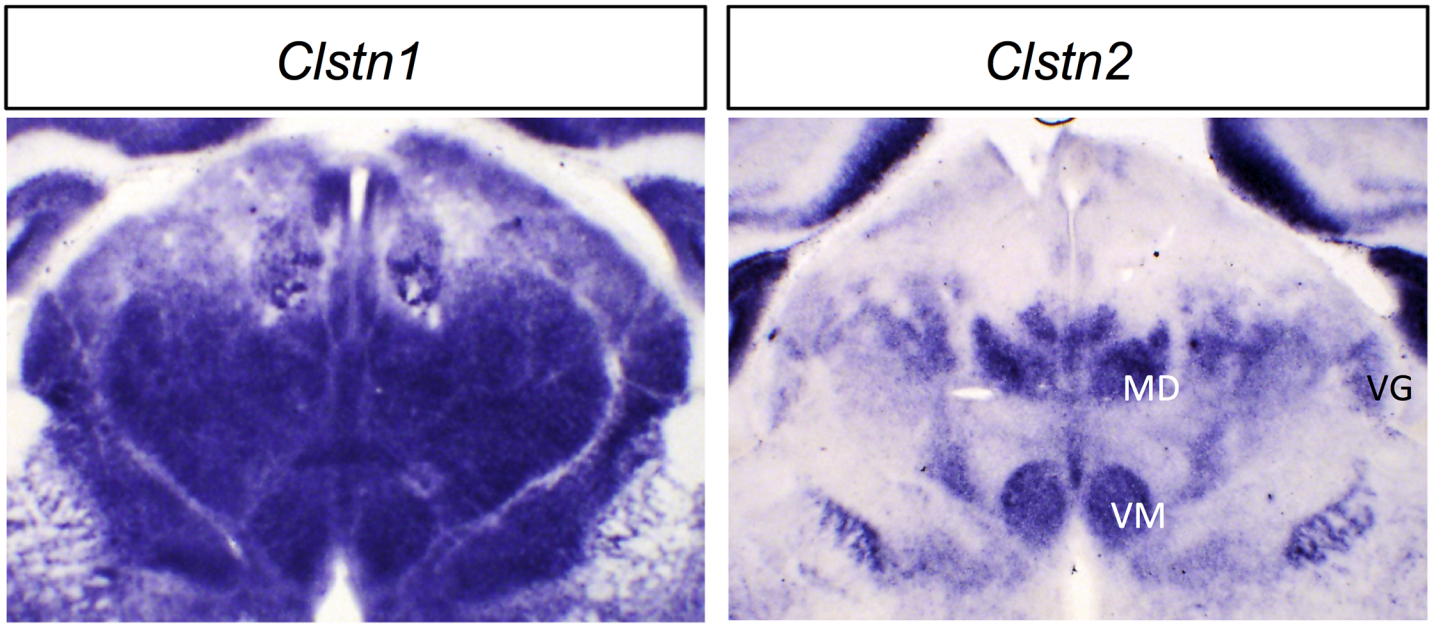

Fig 4. *In situ* hybridization patterns for Calsyntenin family genes in dorsal thalamus at P0. One coronal section for *Clstn1* and *Clstn2* is shown. MD, mediodorsal nucleus; VG, ventral geniculate nucleus; VM, ventromedial nucleus. Scale bar: 500 μm.

https://doi.org/10.1371/journal.pone.0177977.g004

ages of highest interest to us, while higher values gave increasing refinement only at later postnatal stages (S7 Fig).

Gene annotation and enrichment analyses

We used two different approaches to annotate genes, based on known or predicted localisation of the protein product or on known biological function. First, we screened protein products for cellular localisation with an emphasis on the extracellular environment by assessing the outputs of several software tools available online (PSORTII, SignalP, Big-PI Predictor and TMHMM; S4 Table), along with manual curation based on literature sources. The devABA dataset predominantly comprises proteins found in the nucleus (907 proteins), followed by 293 secreted proteins, 286 GPI-anchored or single-pass transmembrane proteins, 267 cytoplasmic (or in organelles) proteins and 240 multi-pass transmembrane proteins (S5 Table Column K). Three genes could not be annotated: *C030002O17Rik*, *LOC433436 and mCG146432*.

Second, we grouped genes by ascribed functions, using GO term annotations and systematic manual curation (S5 Table Column M). The dataset was divided into 6 groups: Group 1: axon guidance pathway and cell adhesion; Group 2: synapse; Group 3: receptor tyrosine kinases and their ligands, and patterning; Group 4: neurotransmission pathway (G-protein-coupled receptors, ion channels, gap junctions); Group 5: chromatin and transcription factor activity; Group 6: other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated.

When a gene fell into more than one group, an order of priority was used (the order in which the groups are presented); for example, a gene encoding a receptor tyrosine kinase that is involved in axon guidance would appear in Group 1 and not in Group 3. Our dataset predominantly included genes involved with chromatin and transcription factor activity (677 genes in Group 5); followed by 360 genes expressing receptor tyrosine kinases and their ligands, and genes involved in patterning (Group 3); 286 genes involved in the axon guidance

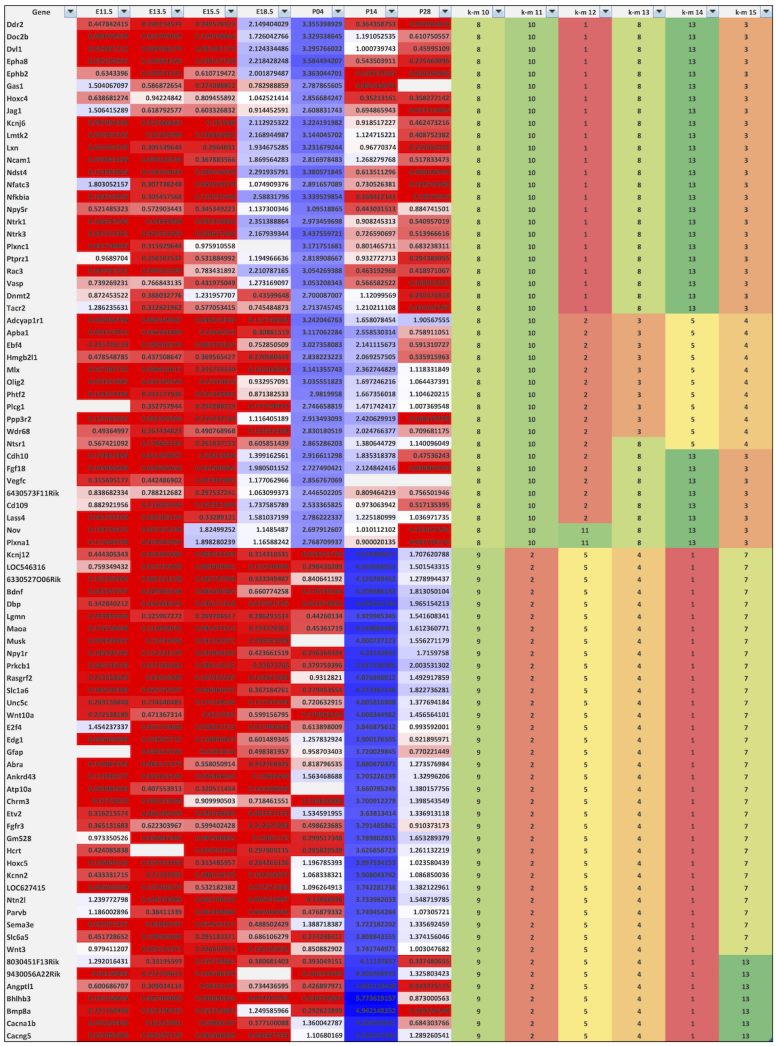

Fig 5. Excerpt of *k*-means clustering analyses. Clustering of a small subset of genes is shown to illustrate the method. Gene names are shown in column 1. Columns 2 to 8 show normalised expression densities per gene per time-point. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Columns 9 to 14 show cluster numbers assigned under $k = 10$ to $k = 15$ clustering, with clusters colour-coded to facilitate visualisation. Clusters are sorted at $k = 10$, followed by $k = 11$, followed by $k = 12$ and so on until $k = 15$. As the input value of k increases, clusters tend to get subdivided as opposed to reshuffled, giving a roughly hierarchical organisation. A clear difference in expression pattern between clusters 8 and 9 at $k = 10$ is evident, with cluster 8 showing peak expression at P4, and cluster 9 showing peak expression at P14. Within cluster 8, additional subclusters become apparent at $k = 12$ and above. The full clustering output for all 1996 genes is shown in S3 Table.

https://doi.org/10.1371/journal.pone.0177977.g005

pathway and cell adhesion (Group 1); 270 in the neurotransmission pathway (Group 4); and, 91 genes involved with synapses (Group 2; S5 Table). Group 6 contained 107 genes involved with or part of the cytoskeleton, extracellular matrix, myelin, signal transduction pathways or expressed metabolic enzymes. This group also included 205 unannotated genes to bring the total to 312 genes.

We hypothesized that different types of proteins would be enriched in clusters with peak expression at time-points relevant to specific developmental functions. In order to test this, we compared the observed numbers for each category in each cluster to the expected value by normal distribution.

k-mean 10	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 7	4.087	0.733	0.318	0.503	0.430	0.478	0.374	131
cluster 6	2.067	0.665	0.465	0.791	0.785	1.423	0.820	268
cluster 5	0.932	2.706	0.674	0.686	0.560	0.946	0.447	106
cluster 2	0.769	0.966	1.548	1.047	1.196	0.769	0.703	163
cluster 4	0.553	0.378	0.426	2.779	1.286	1.003	0.550	157
cluster 0	0.317	0.318	0.421	0.965	1.628	2.069	1.256	384
cluster 8	0.403	0.284	0.303	0.791	3.467	1.070	0.662	192
cluster 9	0.432	0.249	0.226	0.482	0.491	4.237	0.839	214
cluster 3	0.271	0.213	0.190	0.329	0.650	2.949	2.374	253
cluster 1	0.450	0.294	0.225	0.494	0.566	1.175	3.699	128

Fig 6. Summary of expression profiles at $k = 10$. Normalised expression densities were averaged per cluster to see the trends of expression at $k = 10$. Clusters were organised chronologically with those showing early peaks of expression at the top and later peaks of expression at the bottom. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Number of genes per cluster is shown in the rightmost column.

https://doi.org/10.1371/journal.pone.0177977.g006

A variety of trends emerged from these data (Figs 7 and 8). Chi square statistics on the whole dataset showed deviation from expected by random distribution at p-value < 0.001 for either localisation or function (Tables 2 and 3). Further statistics were done on individual categories to establish which ones were enriched in particular clusters. The Bonferroni correction was taken into account due to multiple testing, lowering the threshold to p-value ≤ 0.01 and p-value ≤ 0.008 for localisation and function, respectively. For localisation, the trends observed with secreted and cytoplasmic proteins did not reach statistical significance (p-value > 0.1). However, the distribution of single-pass transmembrane and GPI-anchored proteins, multi-pass transmembrane proteins and nuclear proteins were significant at p-value < 0.0001, p-value < 0.001 and p-value < 0.001, respectively. For functional groups, all trends observed reached statistical significance, at p-value < 0.005 for Group 2, p-value < 0.001 for Group 1, and p-value < 0.0001 for Groups 3 to 6.

By plotting the relative enrichment for each category across clusters, ordered by time-point of highest expression, it was possible to discern what relationships were driving these deviations from a random distribution (Figs 7 and 8), and to assess their biological plausibility. For the localisation data, nuclear proteins were enriched in clusters peaking at early time-points, whereas single-pass transmembrane and GPI-anchored proteins were enriched in clusters peaking from E15.5 to P4, and multi-pass transmembrane proteins in clusters peaking from P4 to P28 (Fig 7). This nicely correlates with the extensive requirement of transcription factors during patterning and differentiation early on (localised in the nucleus), followed by axon guidance molecules expressed extracellularly during axon pathfinding, and finally synaptogenesis and neurotransmission postnatally (including many multi-pass neurotransmitter receptors and ion channels). These inferences were supported by enrichment of functional groups: chromatin and transcription factors early on, axon guidance and synapse molecules in later embryogenesis, and neurotransmission pathway genes at later postnatal stages (Fig 8).

Taken together, these protein localisation and functional enrichment analyses provide a powerful and independent validation of our clustering analyses, indicating that the clusters we have identified with $k = 10$ track valid biological trends. They further suggest that clusters enriched for known molecules involved in axon guidance or synaptic connectivity may also contain novel molecules with these functions. In order to specifically search for such genes, we focused on clusters 2, 4, 0 and 8, which show enrichment for single-pass and GPI-linked proteins and for proteins involved in axon guidance or synaptogenesis. These have peak expression at E15.5 (cluster 2), E18.5 (cluster 4) and P4 (clusters 0 and 8).

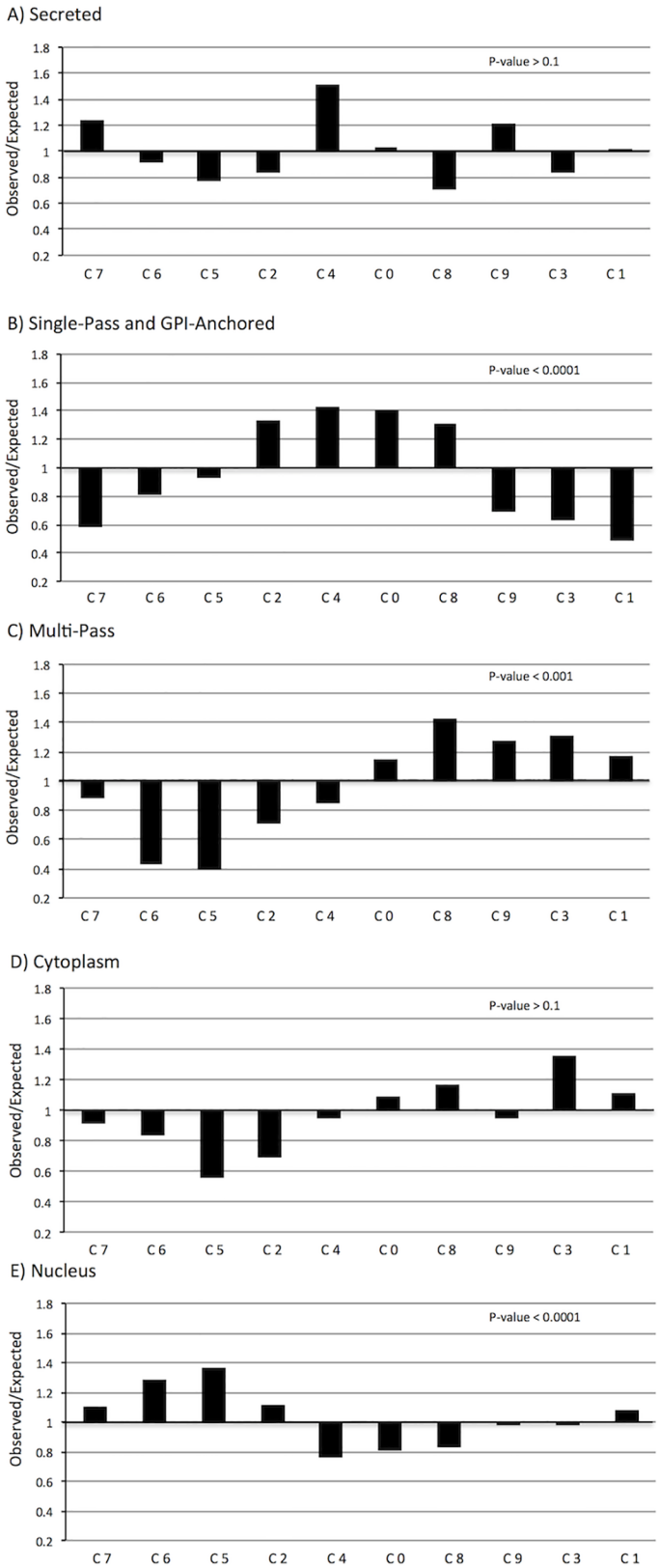

Fig 7. Enrichment of protein localisation categories across clusters. Relative enrichment across clusters is shown by plotting observed/expected numbers of proteins per cluster for each of five groups: secreted

proteins (A), single-pass transmembrane or GPI-anchored proteins (B), multi-pass transmembrane proteins (C), cytoplasmic proteins (D) and nuclear proteins (E). A value of 1 indicates observed data matches expected values, whereas a value below or above 1 indicates decreased or increased counts compared to expected values, respectively. The ten clusters (C0-C9) are organized according to age of peak expression, as in Fig 6. P-values from chi square analyses are shown in upper right corner of each graph. Genes encoding nuclear proteins are enriched in clusters defined by strong early embryonic expression, single-pass transmembrane and GPI-linked protein-encoding genes are enriched in clusters expressed at mid-embryonic stages and multi-pass transmembrane proteins are enriched in clusters expressed at postnatal stages.

https://doi.org/10.1371/journal.pone.0177977.g007

Identifying candidate connectivity labels

Clusters 2, 4, 0 and 8 had a total count of 896 genes of which 426 encode extracellular proteins. We screened through them by visual inspection of the devABA expression data at their peak expression to identify those with selective or differential expression across the thalamus. Genes showing uniform expression were not considered further. We generated a list of 215 genes that potentially encode specific connectivity information (S6 Table). This dataset includes numerous genes already implicated in thalamocortical connectivity (such as *Chl1*, *DCC*, *L1CAM*, *NCAM1*, *Ntn1*, *Robo1* and *Robo2*), which we did not characterize further. However, where previously implicated genes were part of larger families (e.g., cadherins, Ephrins and Eph-receptors, semaphorins), they were included for comparison with other members. Some genes had previously demonstrated roles in axon guidance or synapse formation in other contexts and were of high interest to us, as were those encoding surface proteins not previously implicated in these processes. However, we did not pursue any of the large number of neurotransmitter receptor and ion channel genes, given a lower prior probability of involvement as direct axon guidance or synaptic connectivity labels.

In order to present a summary of the expression patterns of these genes and to enable comparison across genes, we extracted images for each gene from the devABA database of a consistent lateral and/or a more medial or parasagittal section at E18.5 (S8 Fig and Figs 9–20). As the borders between nuclei are still developing at E18.5, it is not possible to definitively match expression patterns to specific nuclei at this stage. These figures are not intended to be comprehensive, as the full dataset can be viewed on the devABA website. They do, however, allow a survey of the general trends of expression patterns across the thalamus and a comparison of multiple members of specific gene families, while also highlighting numerous individual genes with strikingly selective patterns.

In general, the expression patterns were extremely diverse–in fact, we did not detect any two patterns that were identical across these genes. Some genes showed highly selective expression patterns, on in some developing nuclei and off in others. Others showed differential expression, higher in some nuclei than others, with substantial variation in extent of expression across the thalamus. Within developing nuclei, some genes appeared to be expressed at uniform levels, while others showed an uneven or graded distribution. Below, we consider the expression patterns of 73 genes organised by protein family or general function (Table 1 Columns B-C).

Ephrins and Eph-receptors. Twelve members of the Eph and Ephrin gene families show differential or selective expression across the developing thalamus (Fig 9). Multiple members of the *EphA* and *-B* receptor and *Ephrin-A* and *-B* families have been implicated in diverse aspects of thalamic connectivity. The most extensively investigated have been *EphA4*, *EphA7*, *EphrinA2* (*Efna2*) and *EphrinA5* (*Efna5*). Roles for these genes have been demonstrated in parcellation of thalamic nuclei [19], target selection and topography of retinal axons into the thalamus [40], topographic guidance of thalamic axons through the ventral telencephalon [41], areal and layer-specific targeting and topography (reviewed in [42]), reciprocal corticothalamic axon guidance

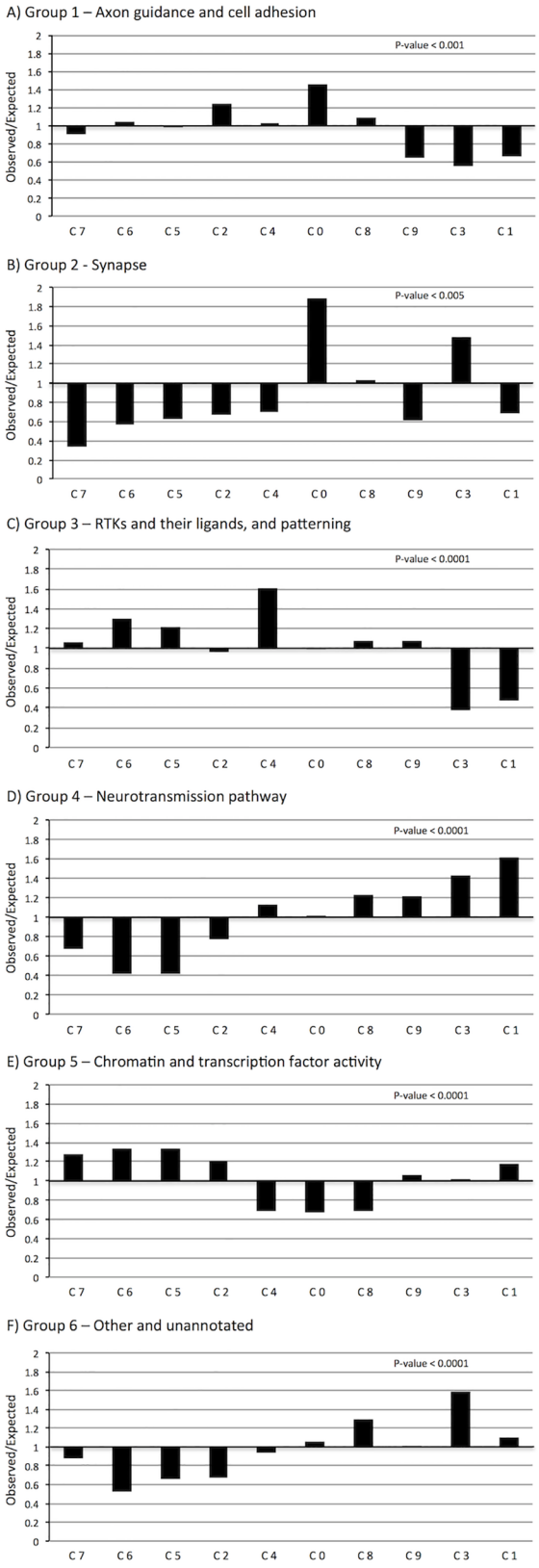

Fig 8. Enrichment of functional groups across clusters. Relative enrichment across clusters is shown by plotting observed/expected numbers of proteins per cluster for each of six groups: Group 1—axon guidance

pathway and cell adhesion (A); Group 2—synapse (B); Group 3—receptor tyrosine kinases and their ligands, and patterning (C); Group 4—neurotransmission pathway (GPCRs, ion channels, gap junctions; D); Group 5—chromatin and transcription factor activity (E); and, Group 6—other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated (F). A value of 1 indicates observed data matches expected values, whereas a value below or above 1 indicates decreased or increased counts compared to expected values, respectively. Clusters 0–9 are organized chronologically. P-values from chi square analyses are shown in upper right corner of each graph. All groups showed statistically significant deviation from expected distributions.

https://doi.org/10.1371/journal.pone.0177977.g008

and target selection [43, 44] and even influences on cortical progenitor proliferation and differentiation dynamics [45]. *EphB1* and *-B2* have also recently been shown to mediate thalamic axon guidance through the ventral telencephalon [46]. The selective expression of several other members in the thalamus, including *EphA3*, *-A5*, *-A6* and *-A8* has also previously been noted [20, 47]. The devABA data confirm these findings and allow a direct comparison of expression patterns across members of these families. In addition, we find that *EphA1*, *EphA10*, *EphB6* and *Efnb3* are also differentially expressed across the developing thalamus.

Cadherins. We report differential or selective expression of twelve members of the cadherin family across the developing thalamus (Fig 10). A number of cadherins have been implicated previously in thalamic connectivity. Takeichi and colleagues showed differential expression of *Cdh6*, *Cdh8* and *Cdh11* in both cortex and thalamus and used dye-tracing to demonstrate that *Cdh6*- or *Cdh8*-positive cortical areas are specifically connected with *Cdh6*- or *Cdh8*-positive thalamic nuclei, respectively [48]. Similarly, but for afferent connections to the thalamus, *Cdh6* has been shown to act as a homophilic targeting molecule for a subset of retinal ganglion cells which project to nuclei mediating non-image-forming visual functions [10]. *N-Cadherin* (*Cdh2*) is required for proper termination of thalamic axons in layer 4 of the cortex [49], and *Cdh2* and *Cdh8* differentially label thalamic terminations in somatosensory cortex barrel center and septal compartments, respectively [50]. Differential expression across thalamic nuclei of the remaining cadherins in this group, *Cdh4*, *7*, *9*, *10*, *12*, *13*, *24* and the atypical cadherin *Fat3*, has not, to our knowledge, been previously reported and greatly expands the possible involvement of this gene family in specifying thalamic connectivity. The diversity of expression patterns is particularly noteworthy; even in this single parasagittal section, no two members of this family show an identical pattern. Given the known potential for heteromeric complex formation between cadherins, these overlapping patterns generate a very large potential combinatorial code.

Protocadherins. A number of members of the unclustered *protocadherin* family show differential expression across the thalamus at E18.5, including *Pcdh1*, *10*, *11X*, *19* and *21* (Fig 11).

Table 2. Enrichment analyses per localisation.

Cluster	Chi Square	Degrees of Freedom	P-Value	Bonferroni Correction	Conclusion
All	122.694	36	< 0.001	No	Enrichment
Secreted	16.328	9	> 0.1	Yes	No Enrichment
Single-Pass and GPI	42.006	9	< 0.0001	Yes	Enrichment
Multi-Pass	31.815	9	< 0.001	Yes	Enrichment
Cytoplasm	13.559	9	> 0.1	Yes	No Enrichment
Nucleus	65.758	9	< 0.0001	Yes	Enrichment

Degrees of freedom: All, (columns #—1) (row #—1) = (5–1) (10–1) = 36; Per Localisation, (c—1) (r—1) = (10–1) (2–1) = 9. Bonferroni correction: $p \leq \alpha$ / # tests = 0.05 / 5 = 0.01

https://doi.org/10.1371/journal.pone.0177977.t002

Table 3. Enrichment analyses per group.

Cluster	Chi Square	Degrees of Freedom	P-Value	Bonferroni Correction	Conclusion
All	206.160	45	< 0.001	No	Enrichment
Group 1	30.770	9	< 0.001	Yes	Enrichment
Group 2	26.554	9	< 0.005	Yes	Enrichment
Group 3	46.368	9	< 0.0001	Yes	Enrichment
Group 4	41.989	9	< 0.0001	Yes	Enrichment
Group 5	73.622	9	< 0.0001	Yes	Enrichment
Group 6	37.325	9	< 0.0001	Yes	Enrichment

Degrees of freedom: All, (columns #—1) (row #—1) = (6–1) (10–1) = 45; Per Group,(c—1) (r—1) = (10–1) (2–1) = 9. Bonferroni correction: $p \leq \alpha / \#$ tests = 0.05 / 6 = 0.008

https://doi.org/10.1371/journal.pone.0177977.t003

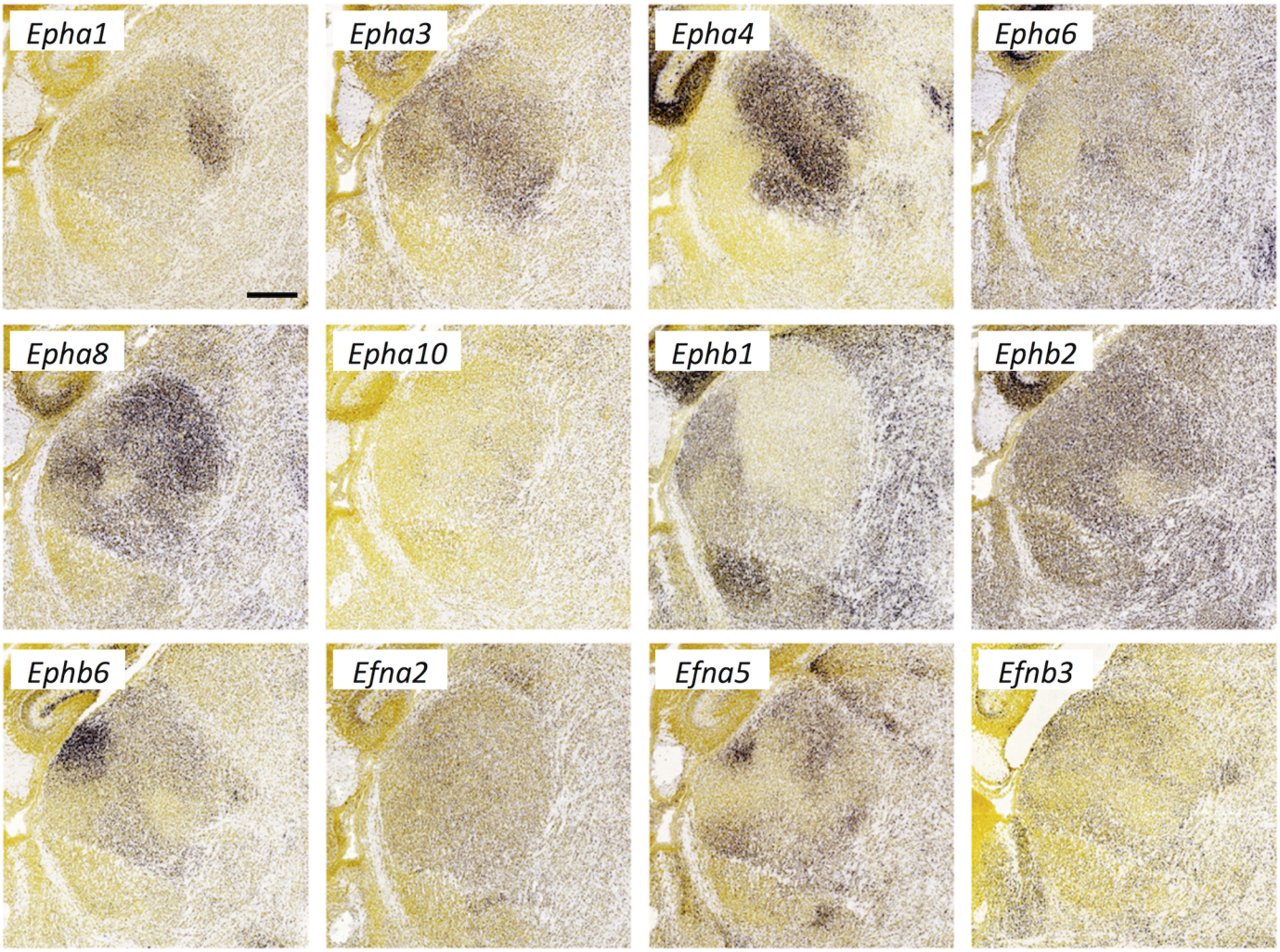

Fig 9. Ephrin and Eph-receptor expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Eph-receptor* (*Eph*) *A1*, *A3*, *A4*, *A6*, *A8*, *A10*, *B1*, *B2*, *B6*, and *Ephrin* (*Efn*) *A2*, *A5* and *B3*. See text for details. A representative and equivalent section of medial thalamus is shown for each gene (see S8 Fig for details on how the sections were selected). Scale bar in *EphA1*: 317 μm.

https://doi.org/10.1371/journal.pone.0177977.g009

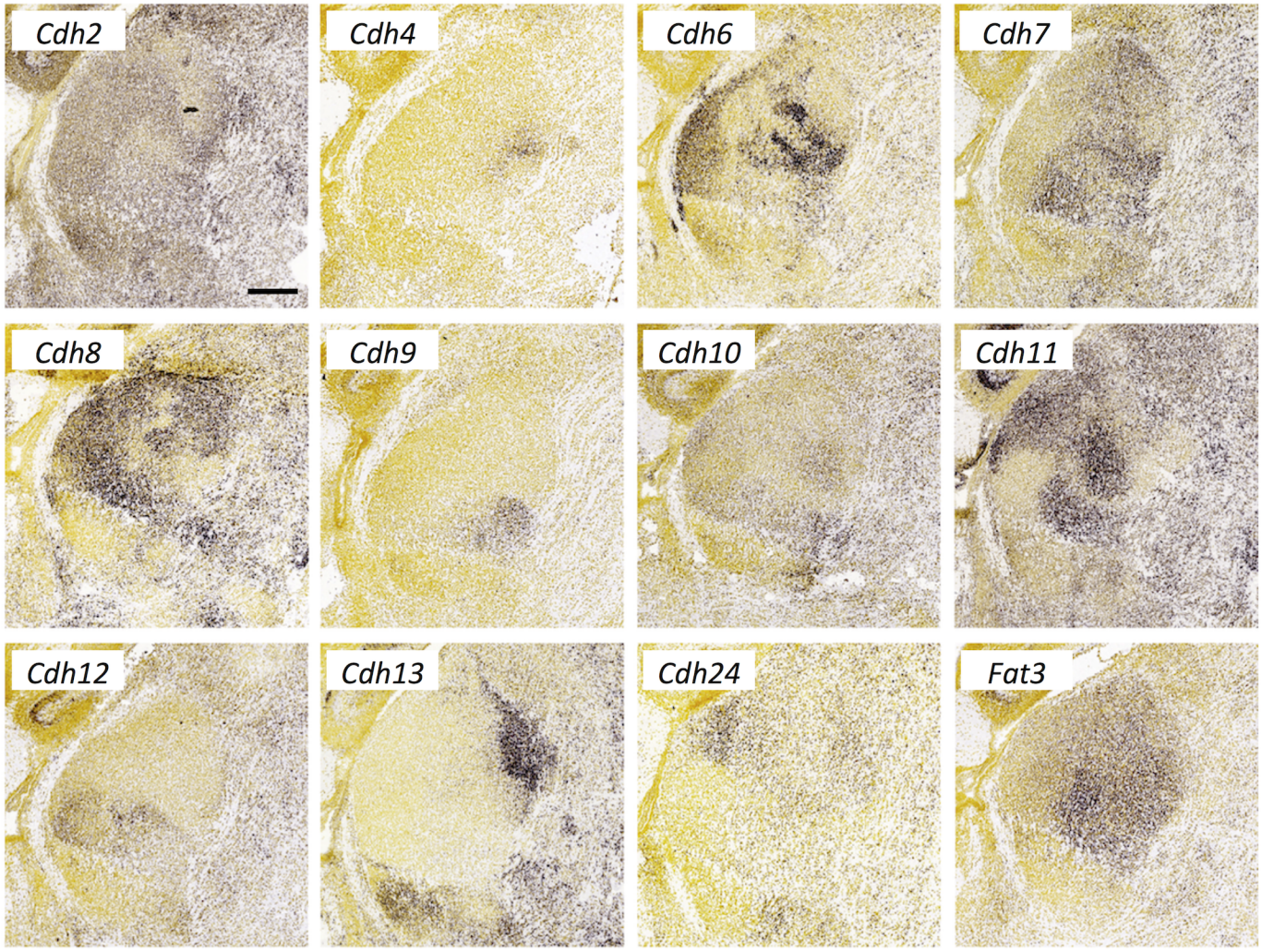

Fig 10. Cadherin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Cadherin* (*Cdh*) *2*, *4*, *6*, *7*, *8*, *9*, *10*, *11*, *12*, *13*, *24* and *Fat3*. See text for details. A representative and equivalent section of medial thalamus is shown for each cadherin (see S8 Fig for details on how the sections were selected). Scale bar in *Cdh2*: 351 μm.

https://doi.org/10.1371/journal.pone.0177977.g010

We note that *Pcdh18* also shows differential thalamic expression at this stage (not shown) but is not included in our clusters of interest due to an earlier time-point of peak expression. These results replicate previous findings reported by Kim *et al.* (2007), who also noted differential expression across thalamus of additional members of this family that are not in the devABA dataset (*Pcdh7*, *8*, *9*, *15* and *17*) [51]. The functional importance of these genes in thalamic connectivity has not been directly tested. Though defects in thalamocortical axon projections occur in *Pcdh10*$^{-/-}$ mice, they have been attributed to functions in patterning of the ventral telencephalon and not to activity in thalamic axons themselves [52]. Many of the protocadherins are also expressed in selective patterns across cortical areas and layers [51], suggesting a possible interaction code between interconnected regions.

Semaphorins and plexins. Among the members of the semaphorin and plexin gene families in the devABA dataset, the ones showing the most selective or differential expression across the thalamus are: *Sema3F*, *Sema6A*, *Sema7A*, *PlxnA1*, *PlxnA2* and *PlxnC1* (Figs 12 and 13).

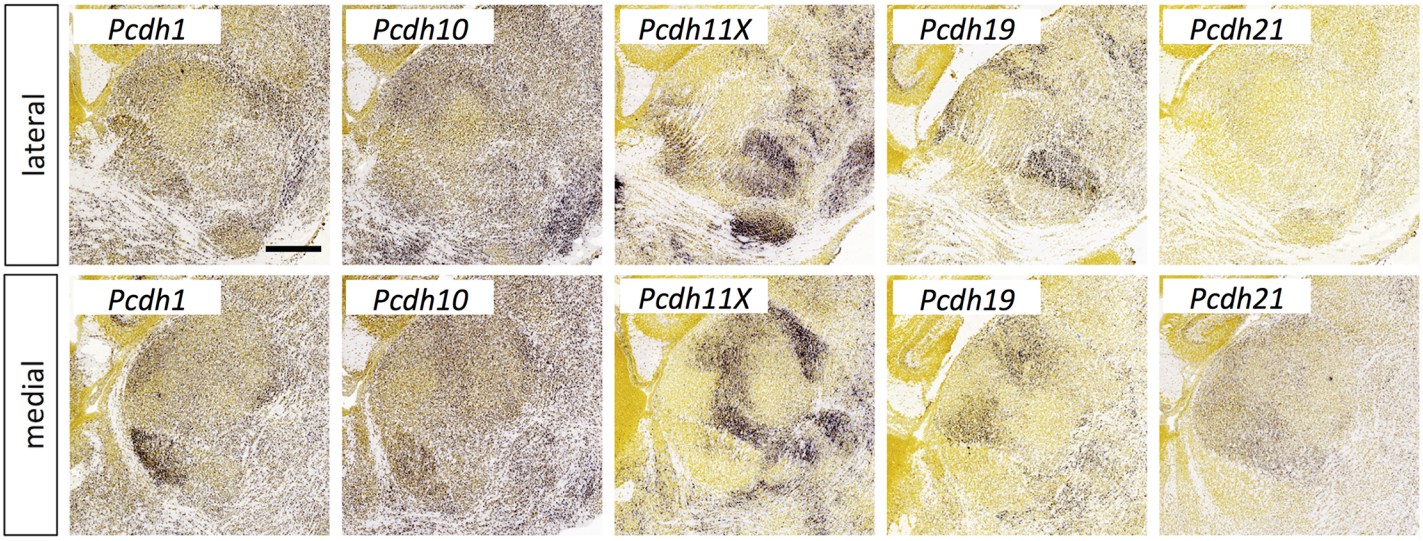

Fig 11. Protocadherin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Protocadherin* (*Pcdh*) *1*, *10*, *11X*, *19* and *21*. See text for details. Two representative and equivalent sections are shown for each protocadherin, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *Pcdh1*: 424 μm.

https://doi.org/10.1371/journal.pone.0177977.g011

PlxnB1 and *Sema6D* are also expressed in developing thalamus, but weakly and more uniformly (not shown). *Sema3F* has been previously implicated in thalamic axon guidance through the ventral telencephalon, and *Sema7A* in regulating thalamocortical axon branching in layer 4 of cortex [53, 54]. However, these functions depend on expression of the semaphorins in these other regions, not in the thalamus itself, where differential expression has not been previously reported. Null mutants of *Sema6A* show misrouted thalamocortical axons [55], specifically from the dLGN [56]; this function may depend on expression of Sema6A in the thalamus itself or in the ventral telencephalon. No roles in thalamic axon guidance or connectivity have yet been described for *PlxnA1*, *PlxnA2* or *PlxnC1*. The expression of *Sema7A* and its receptor *PlxnC1* are notably complementary, while *Sema6A* and its binding partner *PlxnA2* are expressed in an overlapping fashion, with *Sema6A* more widespread and *PlxnA2* restricted to medial regions. Neuropilins also act as co-receptors for secreted semaphorins and have been implicated in thalamic axon guidance in response to Sema3F, signaling through Neuropilin-2-NrCAM co-receptor complexes [57] or Sema3A, acting through Neuropilin-1-CHL1 co-receptors [58]. *Npn2* is not in the ABA developmental dataset but *Npn1* is found in cluster 4 and is differentially expressed across thalamic nuclei, as previously described (not shown).

Receptor tyrosine kinases and phosphatases. In addition to the Eph family of receptor tyrosine kinases, we identified a number of other members of this superfamily and one receptor tyrosine phosphatase showing quite selective expression in the thalamus (Fig 14). Kinases Flt3, Kit (cKit) and Ret have well-documented functions in the immune system and in various cancers, while their functions in the nervous system have been explored to varying extents. No functions for Flt3 in the nervous system have been reported. Kit has been shown to have effects on cortical neuronal migration and axonal extension [59], but has not been functionally implicated in thalamic development, though Kit-ligand (Steel) is also highly expressed in developing thalamus [20]. Ret has multiple, well-described roles in development of the enteric nervous system, as a receptor for GDNF, and mutations in this gene are an important cause of Hirschsprung disease. While expression in thalamus has been noted before [60], this has not been

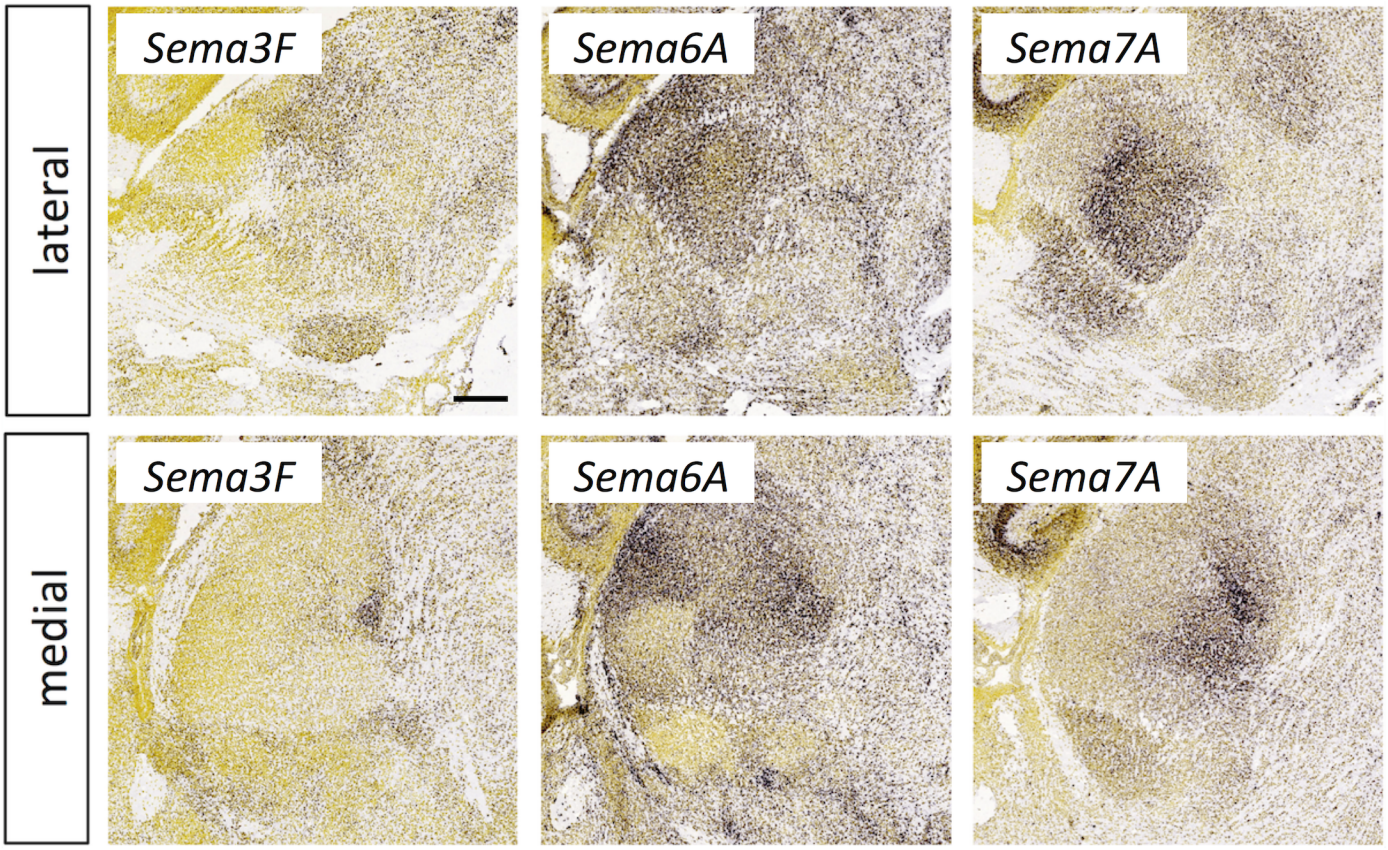

Fig 12. Semaphorin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of semaphorins (*Sema*) *Sema3F*, *Sema6A* and *Sema7A*. See text for details. Two representative and equivalent sections are shown for each semaphorin, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *Sema3F*: 235 μm.

https://doi.org/10.1371/journal.pone.0177977.g012

described in detail nor has a functional role in thalamic development been established. The neurotrophin receptors Ntrk2 (TrkB) and Ntrk3 (TrkC) have been studied in this context, with a demonstration that TrkB in particular is required for normal segregation of thalamic afferents in barrel cortex [61]. Ptpru (also called RPTP-lambda or PTP-RO) is closely related to RPTP-kappa and RPTP-mu and similarly mediates homotypic adhesion [62]. Ptpru has also been shown to associate with c-Kit and to negatively regulate its signaling [63]; however, the expression of these genes in thalamus appears largely non-overlapping (Fig 14). There are no published reports of Ptpru function in the nervous system, though many other members of the RPTP family play important roles in axon guidance and synaptic connectivity [64].

Immunoglobulin superfamily. Multiple Ig superfamily molecules have been implicated in axon guidance or neuronal connectivity in the developing thalamus, including Robo proteins, DCC, various cell adhesion molecules (NCAM, L1, CHL1. . .) and others (reviewed in [9]). Here, we highlight the selective expression of five Ig superfamily members not previously implicated in thalamocortical development (Fig 15). Alcam (also known as CD166, Neurolin, or DM-GRASP/SC1/BEN) plays a role in guidance and fasciculation of motor and retinal axons [65], and in topographic mapping of retinal axons across the superior colliculus, possibly through interaction with adhesion molecule L1 on retinal ganglion cells [66]. Cadm1 (better known as SynCAM) is involved in synapse organisation [67] and axon guidance [68]. Cd47

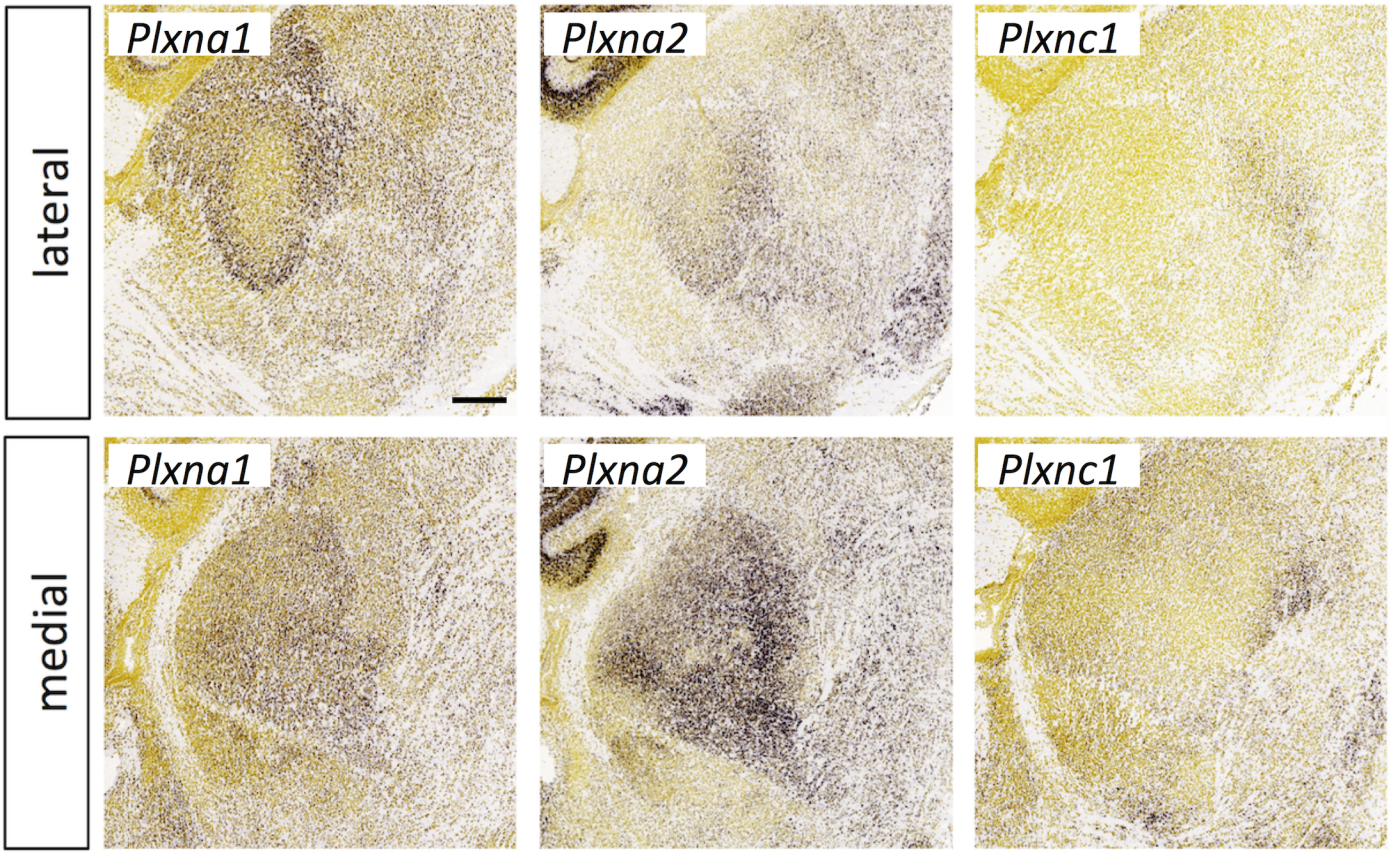

Fig 13. Plexin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of plexins *(Plxn) PlxnA1*, *PlxnA2* and *PlxnC1*. See text for details. Two representative and equivalent sections are shown for each plexin, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *PlxnA1*: 248 μm.

https://doi.org/10.1371/journal.pone.0177977.g013

is a multi-pass transmembrane molecule with a single extracellular Ig domain. It acts as a receptor for presynaptic organising molecule SIRPalpha [69] and has been implicated in various processes in cerebellar development [70]. Cntn6 (NB3) is a member of the contactin family of adhesion molecules, which have widespread roles in neural development [71]. *Cntn6* null mutants have defects in synapse formation in cerebellum [72] and hippocampus [73]. Our results replicate previous reports of expression of *Cntn6* in anterior thalamic nuclei [74, 75]. *Mdga1* is expressed in a layer- and area-specific manner in the cortex [76] and mutation of the gene leads to a delay in cortical neuronal migration [77]. More recently, Mdga1 has also been found to interact with neuroligin-2 to negatively regulate inhibitory synapse formation [78, 79].

Leucine-rich repeat superfamily. Many members of the extracellular leucine-rich repeat (LRR) superfamily have been implicated in axon guidance, synaptic target selection and other aspects of neural development [23, 80]. LRR superfamily genes are highly under-represented in the devABA dataset, with no members of the *Lrfn*, *Slitrk*, *NGL*, *Elfn* or *LINGO* subfamilies and only one representative of each of the *Lrrtm*, *Lrrc*, *Lrrn* and *Amigo* subfamilies. Amongst the LRR genes that are represented, we identified three with selective expression in developing thalamus: *Lgi2*, *Lrrn3* and *Rtn4rl1* (Fig 16). *Lgi2* encodes a secreted LRR protein of the Lgi subfamily that has been implicated in synaptic development and epilepsy [81]. Lgi1 regulates

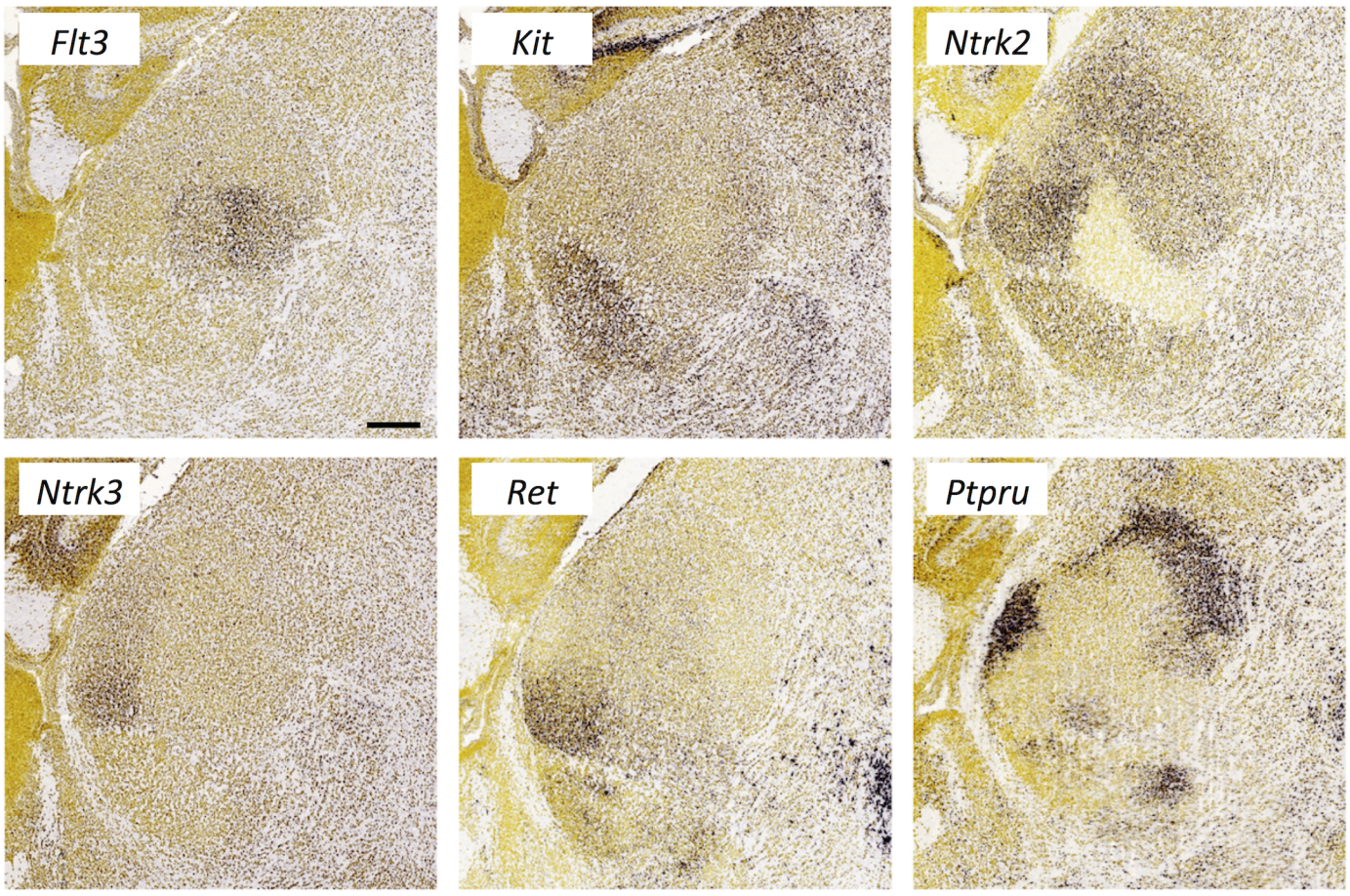

Fig 14. Expression of receptor tyrosine kinases and phosphatases in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Flt3*, *Kit*, *Ntrk2*, *Ntrk3*, *Ret* and *Ptpru*. See text for details. A representative and equivalent section of medial thalamus is shown for each gene (see S8 Fig for details on how the sections were selected). Scale bar in *Flt3*: 216 μm.

https://doi.org/10.1371/journal.pone.0177977.g014

synapse formation and function through interactions with postsynaptic Adam22 and presynaptic Adam23 proteins [82, 83]. Lgi2 likely has similar functions as it also interacts with these Adam proteins and mutations in the gene cause epilepsy in dogs [84]. Lrrn3 (or Nlrr3) is a member of the LRR-Ig-FN3 group [23]. Roles for Lrrn1 and Lrrn2 have been described in hindbrain development in chick [85, 86]; otherwise this family remains poorly characterised. Selective expression in dorsal thalamus has been commented on previously, in particular its complementarity to the expression pattern of *Slitrk6* [87]. *Rtn4rl1* is also known as *Nogo-receptor-3*, *NgR3*. NgR's were first isolated as receptors for the inhibitory myelin protein Nogo-66, but also function as receptors for chondroitin proteoglycans [88] and have been implicated in synaptic development in hippocampus [89].

Small gene families. A number of small gene families were identified with compelling expression patterns in developing thalamus (Figs 17 and 18).

Calsyntenins. The devABA data replicate our own *in situ* hybridization results (Fig 17 and see Fig 4), with widespread, but not ubiquitous expression of *Clstn1* and much more selective expression of *Clstn2*.

Cerebellins. Cerebellins are secreted molecules with synaptogenic activity. They interact with multiple partners including the glutamate receptor subunits GluR**δ**2 and neurexins [90]

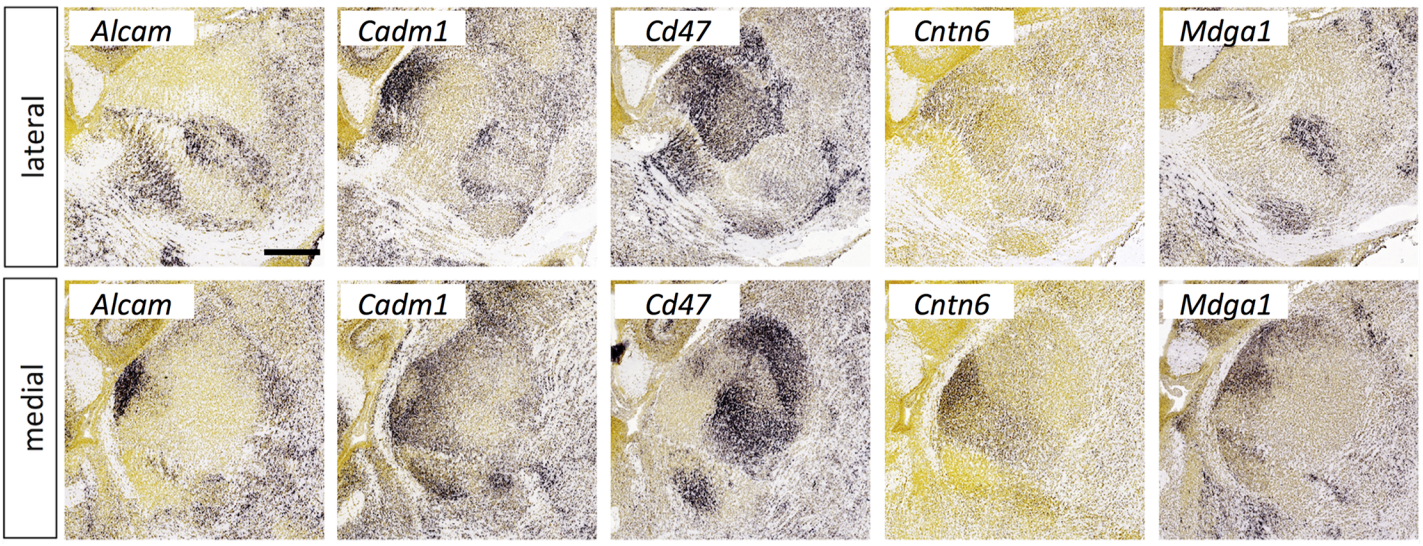

Fig 15. Expression of immunoglobulin superfamily molecules in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Alcam*, *Cadm1*, *Cd47*, *Cntn6* and *Mdga1*. See text for details. Two representative and equivalent sections are shown for each gene, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *Alcam*: 386 μm.

https://doi.org/10.1371/journal.pone.0177977.g015

as well as the netrin receptor DCC [91]. We find strong and selective expression of *Cbln2* and *Cbln4* across thalamic nuclei (Fig 17). Their differential expression across the brain has been documented previously, including in the developing thalamus [92]. By contrast, *Cbln3* is not expressed outside the cerebellum. *Cbln1* is not in the devABA dataset but is also expressed in subsets of thalamic nuclei [92].

Netrin-Gs. Netrin-G1 and 2 are GPI-anchored proteins related to netrins. They act as synaptic cell adhesion molecules via interactions with the transmembrane leucine-rich repeat proteins Netrin-G-ligand (NGL) -1 and -2 [93]. The devABA data match previous reports of broad *Ntng1* expression across the dorsal thalamus and *Ntng2* expression restricted to the habenula (Fig 18) [94]. *Ntng1* expression is not uniform; however; it displays significantly higher expression in some thalamic nuclei than others.

Miscellaneous genes. In addition to the gene families described above, there is a large set of miscellaneous genes encoding surface or secreted proteins with differential expression patterns (Fig 19). These include: *Astn2*, *Cntnap4*, *Dlk1*, *Dner*, *Fzd7*, *Gpc3*, *Lrp8*, *Lypd1* (also known as *Lynx2*), *Odz3* and *Trp53i11* (also known as *Tp53i11* or *PIG11*). These display varying degrees of specificity., For example, *Fzd7* is highly restricted to what may become a single thalamic nucleus, while *Gpc3*, is expressed in a small number of scattered cells throughout the thalamus (and strongly in the reticular nucleus, not shown). *Astn2*, *Cntnap4*, *Dlk1*, *Lypd1* all show quite selective expression in some developing nuclei and not others, while *Dner*, *Lrp8* and *Trp53i11* are expressed more widely, but still differentially. The *Odz3* expression pattern documented in the devABA mirrored that seen in our own *in situ* hybridization experiments extremely faithfully (S9 Fig).

Lrp8 (low-density lipoprotein receptor 8) has been previously directly implicated in specifying thalamic connectivity. Mutation of *Lrp8*, along with the related gene *Vldlr* (very low-density lipoprotein receptor), alters retinogeniculate innervation, through mechanisms that appear independent of the function of these proteins as Reelin receptors [11]. Mutation of *Cntnap4* affects synaptic outputs from GABAergic interneurons and ventral tegmental area

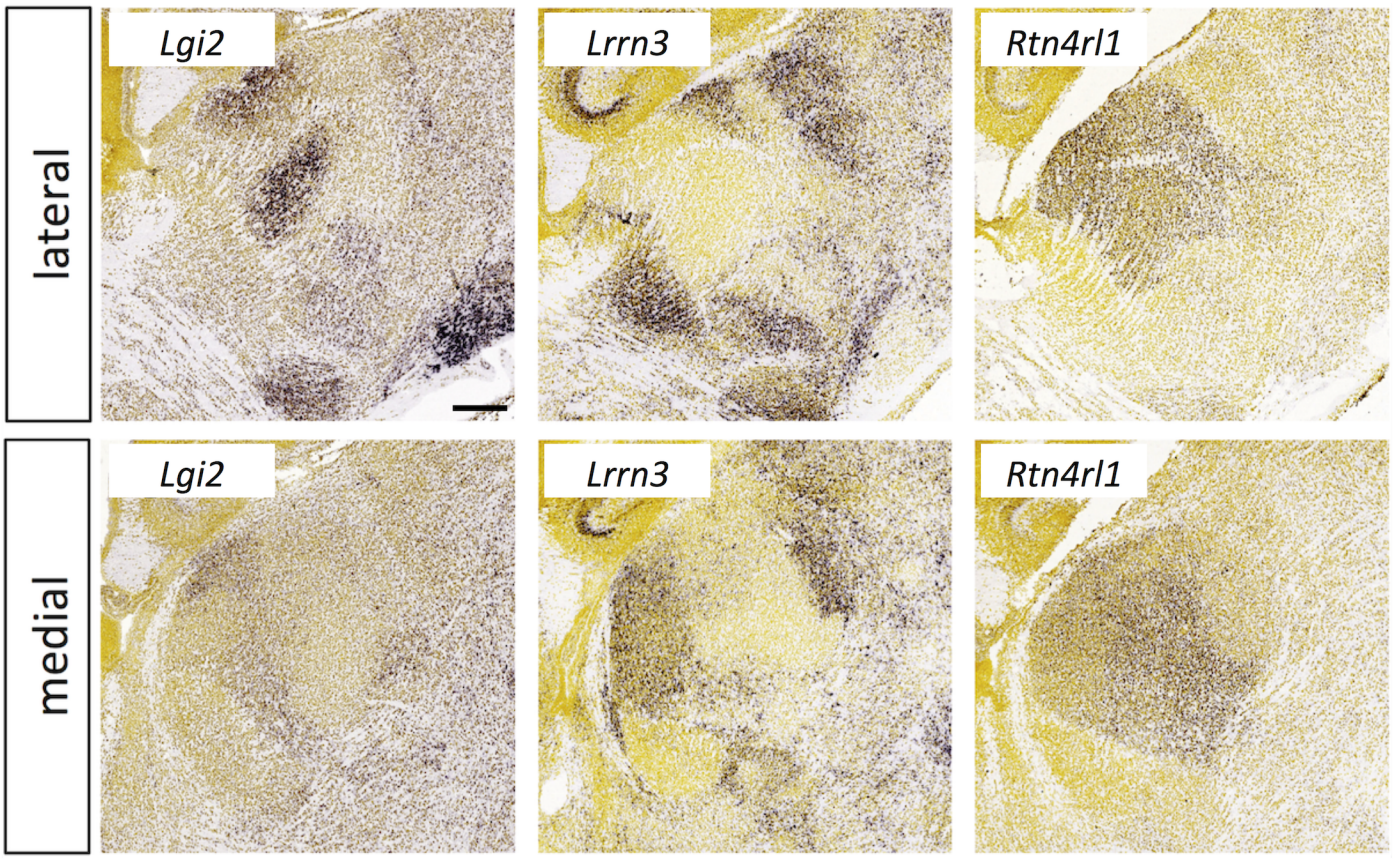

Fig 16. Expression of leucine-rich repeat superfamily members in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Lgi2*, *Lrrn3* and *Rtn4rl1*. See text for details. Two representative and equivalent sections are shown for each gene, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *Lgi2*: 256 μm.

https://doi.org/10.1371/journal.pone.0177977.g016

dopaminergic neurons [95]. Gpc3 (Glypican-3) has been shown to bind to the synaptogenic molecule LRRTM4 [96] and, by analogy, with its close relative Gpc4, may actively specify synapse formation. Fzd7 (Frizzled7) has not been directly implicated in synaptic connectivity but many other members of the Frizzled family have been [97], including Fzd5, which is required for the synapse-organising activity of Wnt7a [98]. *Fzd5* is extremely selectively expressed in the parafascicular nucleus of thalamus [99], which we also find here (data not shown).

A variety of other functions in neural development or function have been shown for *Astn2* (Astrotactin-2) [100], *Dlk1* (Delta-like-1 homolog) [101], *Dner* (Delta/Notch-Like EGF Repeat Containing) [102] and *Lypd1/Lynx2* (LY6/PLAUR Domain Containing 1) [103] but these genes have not been directly implicated in axon guidance or synaptic connectivity. The functions of *Trp53i11* remain largely unknown.

Growth factors and receptors. A variety of growth factors and receptors show selective or differential expression across the late embryonic thalamus (Fig 20). These include: *Bmp3* (Bone morphogenetic protein-3), *Igfbp5* (Insulin-like growth factor binding protein 5), *Inhba* (Inhibin/Activin, beta A subunit), *Gfra1* and *Gfra2* (GDNF Family Receptor Alpha 1 and 2), *Nrn* (Neuritin or Neuritin-1/Nrn1), *Tgfb2* (Transforming growth factor, beta 2), *Vgf* (VGF nerve growth factor inducible) and *Wif1* (WNT inhibitory factor 1).

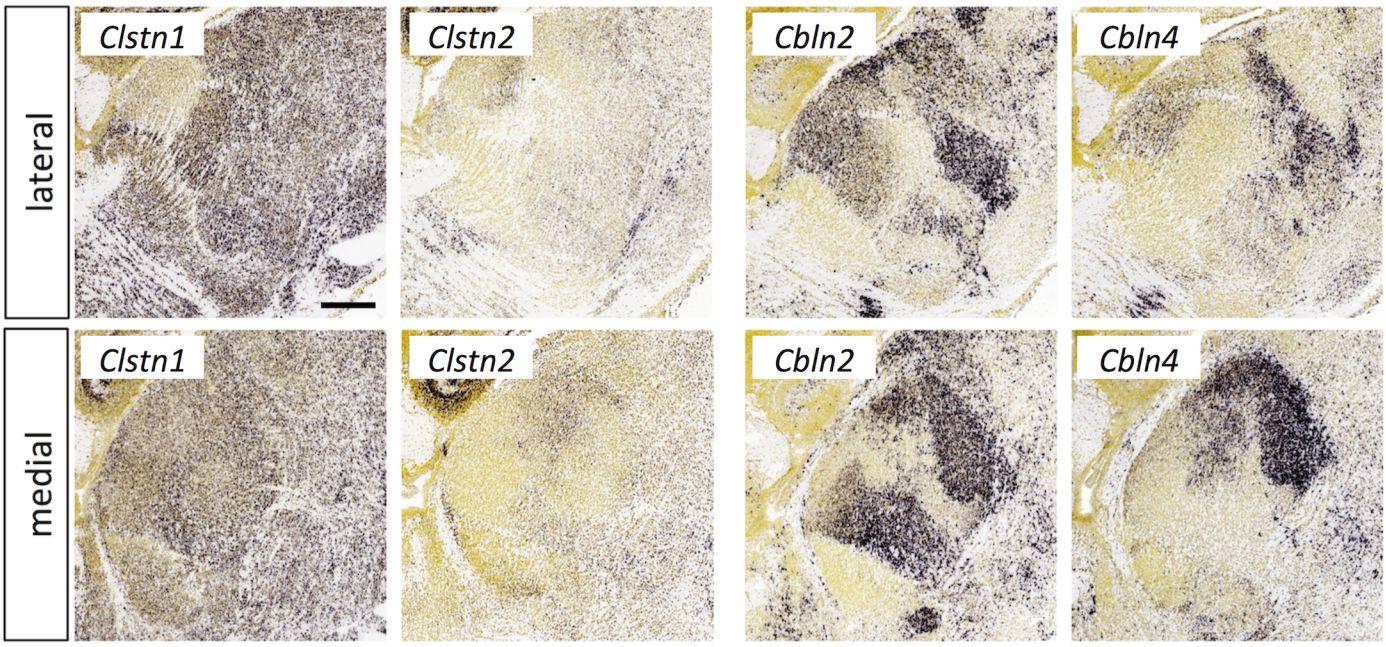

Fig 17. Calsyntenin and cerebellin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Calsyntenin* (*Clstn*) *1* and *2*, and *Cerebellin* (*Cbln*) *2* and *4*. See text for details. Two representative and equivalent sections are shown for each gene, one lateral and the other medial (see S8 Fig for details on how the sections were selected). Scale bar in *Clstn1*: 332 μm.

https://doi.org/10.1371/journal.pone.0177977.g017

Differential expression of the GDNF receptors Gfra1, Gfra2 and Ret has been reported previously in postnatal rat thalamus, with notable overlap in the reticular nucleus [104], also seen in embryonic mouse [105]. These proteins have well defined roles in enteric nervous system formation but their functions in developing thalamus have not been elucidated. Selective expression of Neuritin-1 and VGF across thalamic nuclei has also been noted previously and these secreted molecules may play a role in regulating cortical differentiation in areas innervated by these thalamic axons [106]. Igfbp5, Inhba, Tgfb2 and Wif1 all have known roles in differentiation, proliferation and apoptosis in the developing brain but have not been directly implicated in axon guidance or synaptic connectivity. There are no described roles for Bmp3 in nervous system development.

Discussion

Our dual screening approach has identified 82 genes encoding candidate connectivity labels in the developing thalamus. Several lines of evidence make these molecules highly plausible candidates to be involved in specifying some aspects of thalamic connectivity: (i) they encode proteins with motifs that are common in known connectivity molecules; (ii) some of them are members of gene families with already known functions in axon guidance or synaptic connectivity; (iii) they cluster with known axon guidance or synaptogenesis molecules based on their temporal pattern of expression in the embryonic thalamus; and (iv) they are selectively or differentially expressed across dorsal thalamic nuclei at stages when connectivity is being established.

The genes we have identified include multiple members of large families where some genes have been previously implicated, such as Ephrins and Eph receptors, cadherins and protocadherins, semaphorins and plexins, and Odz/teneurin genes, where we have expanded the

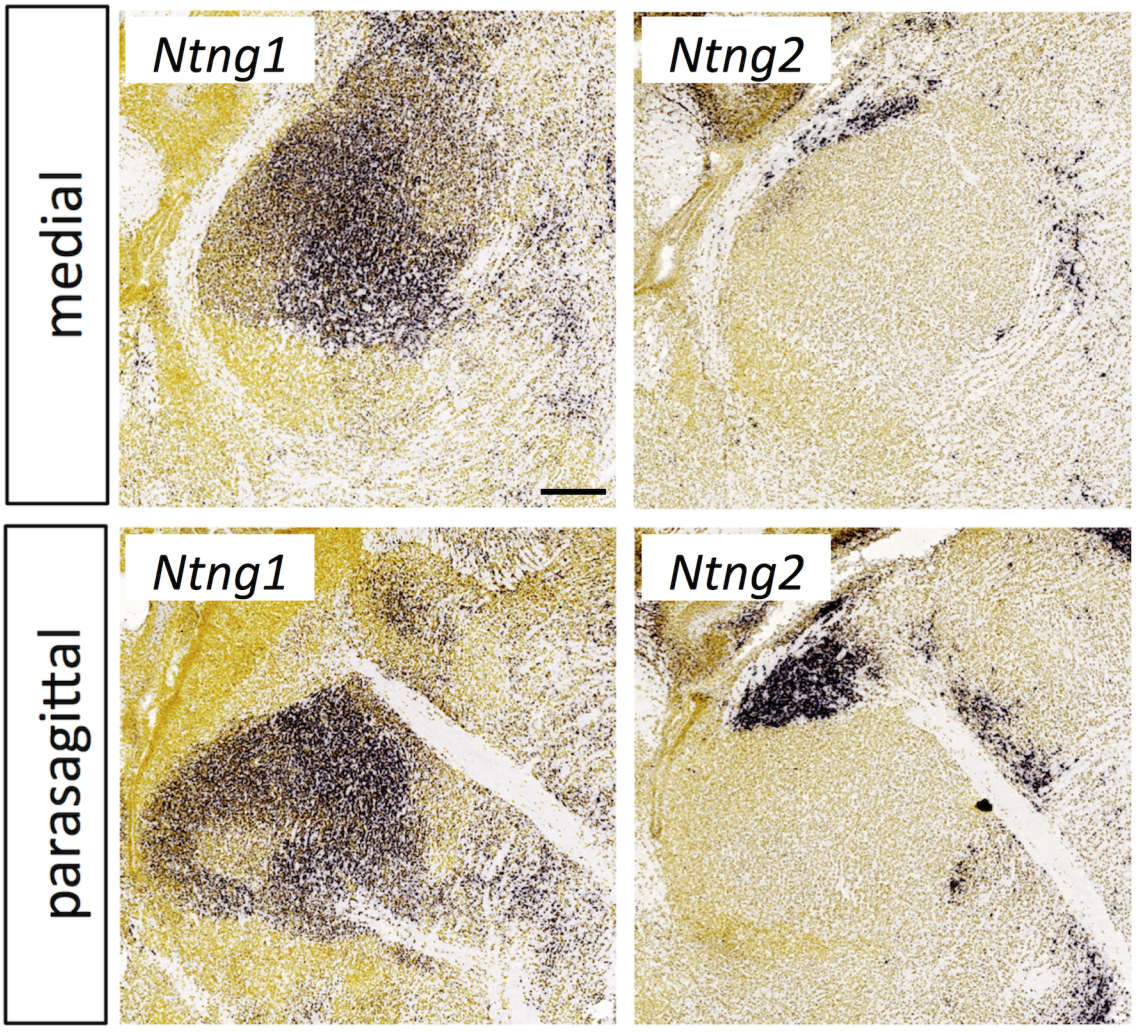

Fig 18. Netrin-G expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Netrin-G* (*Ntng*) *1* and *2*. See text for details. Two representative and equivalent sections are shown for each netrin-G, one medial and the other adjacent medially (see S8 Fig for details on how the sections were selected). Scale bar in *Ntng1*: 255 μm.

https://doi.org/10.1371/journal.pone.0177977.g018

number of candidates potentially involved. Our findings also highlight the selective expression in thalamus of a number of genes or gene families recently implicated in synaptic connectivity in other regions, including several Ig and LRR family members, calsyntenins, cerebellins and Netrin-Gs. In addition, we have found strikingly selective expression of many genes from diverse families not previously implicated in either thalamocortical development or neuronal connectivity.

The systematic nature of the devABA data provides the opportunity to collate and compare these expression patterns across gene families, and to draw some general inferences about the molecular logic specifying thalamic connectivity. The most striking aspect of these patterns is how unique they all are. Across these 82 genes, even when considering only one or two anatomical sections, we did not observe any two patterns that we would say are identical. This clearly did not have to be the case. There might well have been trends or consistent sub-patterns that many genes fell into. Most of the differentiation of expression levels (at E15.5 or

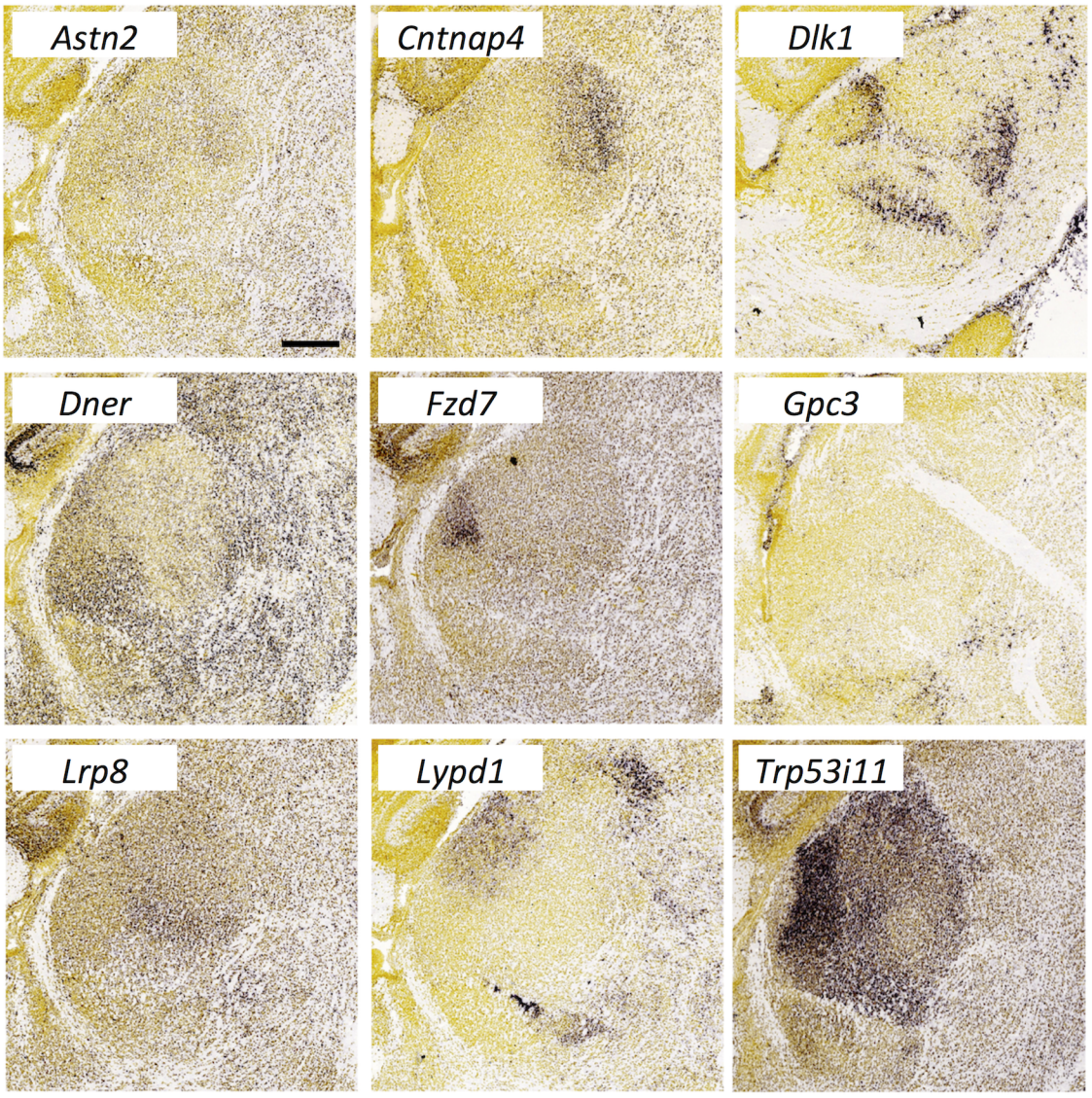

Fig 19. Expression of genes encoding miscellaneous surface or secreted molecules in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Astrotactin2 (*Astn2*), Contactin-associated protein-4 (*Cntnap4)*, Delta-like-1 homolog (*Dlk1*), Delta/Notch-Like EGF Repeat Containing (*Dner*), Frizzled7 (*Fzd7*), Glypican-3 (*Gpc3*), Low-density lipoprotein receptor 8 (*Lrp8*), LY6/PLAUR Domain Containing 1 (*Lypd1*) and *Trp53i11*. See text for details. A representative and equivalent section is shown for each molecule (see S8 Fig for details on how the sections were selected). *Dlk1* expression is shown laterally, while the other gene expression patterns are shown medially with *Gpc3* more medial than the rest (parasagittal). Scale bar in *Astn2*: 314 μm.

https://doi.org/10.1371/journal.pone.0177977.g019

E18.5) reflects discrete protonuclear divisions, rather than broad gradients across the whole dorsal thalamus, which, a priori, was just as likely an outcome. This indicates that thalamic nuclei are highly differentiated from each other, with each one displaying a unique repertoire of these molecules. There is, in addition, clear evidence of non-uniform gene expression within nuclei, reflecting the known sub-differentiation of these fields. Thus, while initial guidance of thalamic axons may be defined by general topography (reviewed in [9]), later connectivity decisions are likely regulated by combinations of more spatially discrete labels.

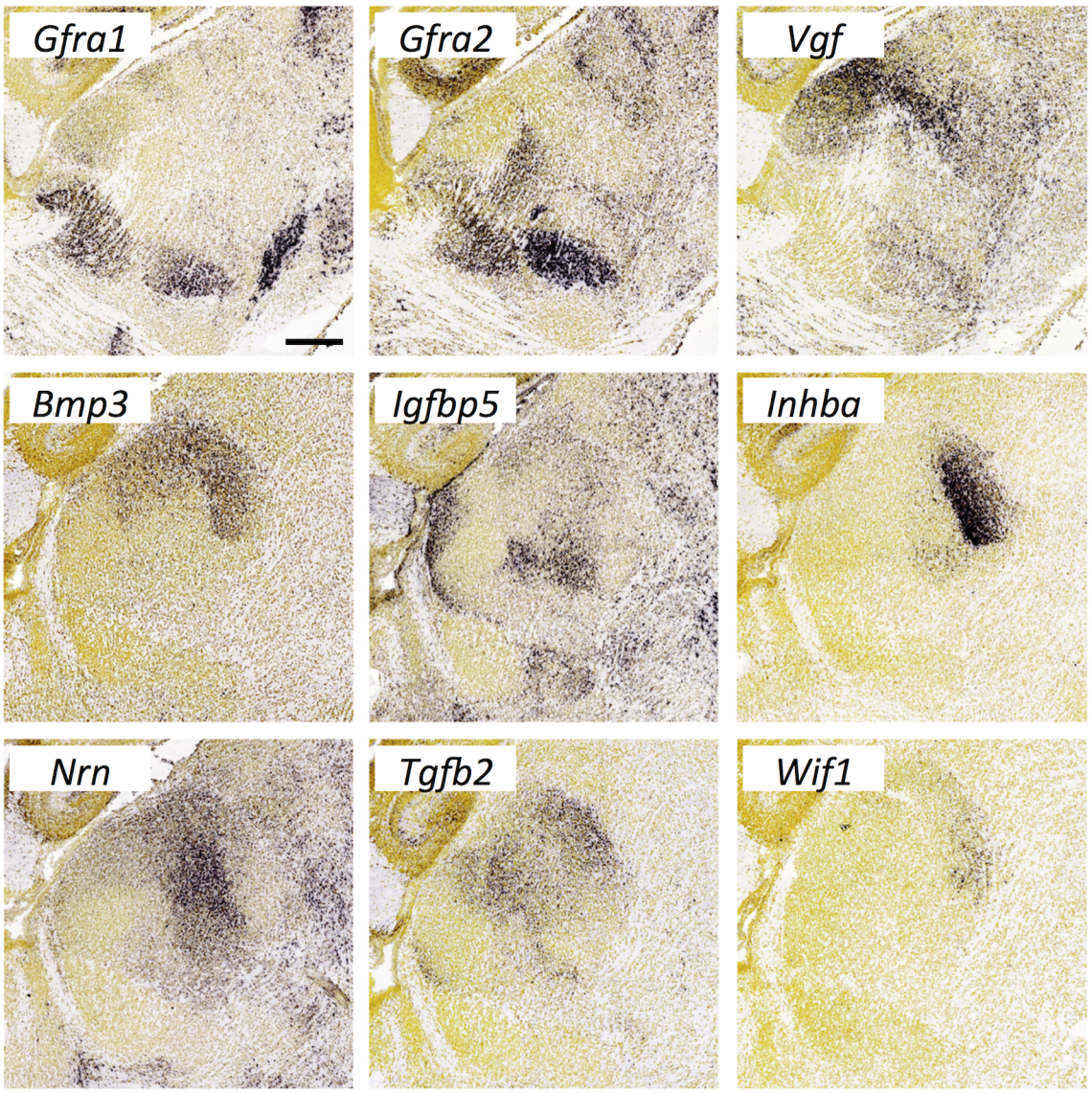

Fig 20. Expression of growth factors and receptors in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of *Gfra1* and *Gfra2* (GDNF Family Receptor Alpha 1 and 2), *Vgf* (VGF nerve growth factor inducible), *Bmp3* (Bone morphogenetic protein-3), *Igfbp5* (Insulin-like growth factor binding protein 5), *Inhba* (Inhibin/Activin, beta A subunit), *Nrn* (Neuritin or Neuritin-1/Nrn1), *Tgfb2* (Transforming growth factor, beta 2) and *Wif1* (WNT inhibitory factor 1). See text for details. A representative section is shown for each gene (see S8 Fig for details on how the sections were selected). *Gfra1*, *Gfra2* and *Vgf* expression is shown laterally, while the other gene expression patterns are shown medially. Scale bar in *Gfra1*: 299 μm.

https://doi.org/10.1371/journal.pone.0177977.g020

There is indeed strong evidence for combinatorial interactions, both within and between the gene families identified. For example, co-expression of Ephrins and Eph-receptors in cis can alter responses to these proteins in trans [107]. Similar modulatory interactions in cis have been observed for Class 6 Semaphorins and Plexin-A proteins, thus generating combinatorial functional diversity [108–110]. Cadherins and protocadherins can also form heteromeric complexes, either within or across these two subfamilies [111, 112], which can alter function, as with a cis interaction between Pcdh19 and Cdh2, which generates a novel trans-adhesive complex [113]. Ig superfamily members of the L1 and Cntn families can mediate homophilic

adhesion, but also modulate signaling of semaphorin and Ephrin pathways in cis [71, 114], while Cntns and Cntnaps also form functional complexes that regulate connectivity (reviewed in [115, 116]). Multiple proteins in our dataset also interact with Neuroligin-Neurexin complexes implicated in synaptic development, including cerebellins [90], calsyntenins [39], Igsf9b [25], Mdga1 [78, 79] and Gpc3 [96]. This pathway may thus represent a convergence point for combinatorial functions of multiple regulators of synaptogenesis, also including diverse members of the LRR superfamily and LAR protein tyrosine phosphatases not surveyed here (reviewed in [115, 116]).

Whether the genes we have identified actually encode connectivity labels will, of course, require functional experiments, but the list here is at least likely highly enriched for such molecules. It is by no means comprehensive, however, as both of our strategies had important limitations. The first approach concentrated on genes that are conserved from vertebrates to invertebrates, based on the simple rationale that many of the major families of axon guidance molecules are conserved. Such molecules may be expected to play more important and thus more obvious roles in nervous system development than newly evolved ones, with concomitantly stronger phenotypes when mutated. In addition, functional analyses of such gene families may be further simplified in invertebrate model systems, which often only have one gene copy, whereas mammals often have multiple genes with possibly redundant functions. Thus, while this strategy identified plausible genes that are highly amenable to functional analyses, it excluded the large fraction of the mammalian proteome that is not obviously conserved in invertebrates.

Our second strategy took advantage of the Allen Brain Atlas database of expression in the developing mouse brain [117]. This database enabled a systematic and quantitative comparison of expression profiles across seven embryonic and postnatal stages, which highlighted several clusters enriched for known or putative connectivity labels. We integrated annotations on protein topology, structural motifs and localisation to better characterise the genes in this dataset and enable us to recognise possible surface labels. However, the dataset is limited to 2002 genes, which were selected on the basis of functional relevance to brain development and neurodevelopmental disorders, and which fall into several categories: (i) transcription factors, (ii) known regional or cell-type-specific markers, (iii) neurotransmitters and receptors, (iv) genes associated with patterning or axon guidance signaling pathways and (v) highly studied common drug targets, including GPCRs, ion channels, cell adhesion molecules and genes associated with neurodevelopmental disorders. While this set is thus itself likely enriched for connectivity labels, it is also certain that many others are not represented. For example, extracellular leucine-rich repeat (LRR) proteins are a major class of neuronal connectivity labels [118], with 135 such genes in the mouse genome [23], but only a small number are included in the devABA dataset.

In summary, the extensive set of genes identified here as candidate connectivity labels provides a strong starting point and context for functional analyses of single molecules and of their potential combinatorial interactions, which may coordinately specify the complex connectivity patterns of thalamic nuclei.

Materials and methods

Ethical approval

All animal work was approved by the Ethics Committee of Trinity College Dublin and the animal licence to KJM (Ref. B100/3527) from the Department of Health and Children. Newborn mice were sacrificed by decapitation.

Mouse *in situ* hybridization

Digoxigenin (DIG)-labeled antisense cRNA probes for *in situ* hybridization were designed to encompass a section of coding sequence and 3'UTR >500bp in length. Briefly this involved TA cloning of PCR products into the TOPO vector (Invitrogen) and simultaneous synthesis and DIG-labeling of RNA transcripts from linearised vector using T7- or Sp6- RNA polymerase. Some cRNA probes were transcribed directly from the PCR products amplified with a sense primer and an antisense primer with T7 promoter sequence at its 5'-end. Primers used were:

Kirrel1_for TGACTATGGCACAGAGCCTA
Kirrel1_rev GAAGGAGAGGAAAGGGCA
Kirrel2_for CAAAAGAACTTGGTCCGGAT
Kirrel2_rev CACTTGGAGAGATGGATTCAC
Kirrel3_for TACGTGCAGTTTGACAAGGC
Kirrel3_rev AGGCTGCAAGGAATACAGAC
Igsf9_for AGACTCACCTCCTGCAAATC
Igsf9_T7Rev GCGGTAATACGACTCACTATAGGGCATTCCTGTTCAGCTCCCA
Igsf9b_for TGACACATTCCACAACGG
Igsf9b_T7Rev GCGGTAATACGACTCACTATAGGGCCCTTATTCCACTTCACCACAG
Sdk1_for CATCATACACTGTGGACCTG
Sdk1_T7rev GCGGTAATACGACTCACTATAGGGAGAAGTTCTGCACCCGTC
Sdk2_for AATGGTGGTTCTTAGTGGTCA
Sdk2_T7rev GCGGTAATACGACTCACTATAGGGCAAACGAAGGATGAGAACC
Odz1_for CATTCAGTGTGAGCTCCAGA
Odz1_T7rev GCGGTAATACGACTCACTATAGGGCTGACGCAAAGGCAGAGAT
Odz1_2_for TGAGGTCCAATATGAGATCC
Odz1_2_T7rev GCGGTAATACGACTCACTATAGGGCAACATCATAATCTCTTTGCCC
Odz2_for AGAGTCAAGCAAGCGAGAA
Odz2_T7rev GCGGTAATACGACTCACTATAGGGCGAGGTGGCAGCAGGTTAT
Odz2_2_for GTGCAATATGAGATGTTCCG
Odz2_2_T7rev GCGGTAATACGACTCACTATAGGGCAGTAAAGTGGACGAGCTTGG
Odz3_for GAAGAGTCAACAGTGGGAAGA
Odz3_T7rev GCGGTAATACGACTCACTATAGGGCTGCTACAGGAGAATCTGCAC
Odz3_2_for CAACCAAATCATTTCCACG
Odz3_2_T7rev GCGGTAATACGACTCACTATAGGGCTGGATTAGTTTGGTGAGCG
Odz4_for TTCAGAAGCAACTCAAGGC
Odz4_T7rev GCGGTAATACGACTCACTATAGGGCCAACACAGTCAGGAATACGG
Odz4_2_for AGAAGGAGCTGAAGGTGG
Odz4_2_T7rev GCGGTAATACGACTCACTATAGGGCCACGATGCTGTTGCTACTC
Neto1_for GGCTCCAGAACTGTGTATATCC
Neto1_T7rev GCGGTAATACGACTCACTATAGGGCGAGTCTTGCCGAAGGAATA
Neto1_2_for AGAAGTCAGTGCAGTGTGG
Neto1_2_T7rev GCGGTAATACGACTCACTATAGGGCGAGCTGCTTCCGTCATAG
Neto2_for CATCAGGAATTGTCTTGGTC
Neto2_T7rev GCGGTAATACGACTCACTATAGGGCCTGCTCTGCCTGACTTAACA
Neto2_2_for GAATCAAGCACATTCCTGC
Neto2_2_T7rev GCGGTAATACGACTCACTATAGGGCGGATCACTCCGACTCCTG
Clstn1_for CACCCTGAACATCGATCC
Clstn1_T7rev GCGGTAATACGACTCACTATAGGGCTCACCCTCTGTCGACAAG

Clstn2_for GCCACTGTCGTCATCATTATC
Clstn2_T7rev GCGGTAATACGACTCACTATAGGGCATGGTGACATCATCTCCAGA
Clstn3_for TCCCAGTCATGTGCTCAGT
Clstn3_T7rev GCGGTAATACGACTCACTATAGGGCAGTAGAGAAGGACACAGCAGC

In situ hybridization (ISH) was carried out on vibratome-sectioned C57BL6 mouse brains (Jackson Laboratories) as described in [119] with modifications. To obtain brains at E15, timed pregnancy mice were sacrificed by cervical dislocation and embryos were fixed and stored in 4% paraformaldehyde (PFA)-PBS at 4˚C until use. For P0, brains were dissected out prior to immersion in 4% PFA-PBS -at 4˚C. P9 brains were fixed by perfusion followed by immersion in 4% PFA-PBS at 4˚C. Fixed brains were embedded in 3% agarose and 70-μm sections were obtained on a vibratome (VT1000S Leica). Sections were washed twice in PBST (PBS containing 0.1% Tween-20), permeabilised in RIPA buffer (150mM NaCl, 50mM Tris-HCl pH 8.0, 1mM EDTA, 1% Nonidet-P40, 0.5% sodium deoxycholate, 0.1% SDS) and post-fixed in PBS containing 4% PFA and 0.2% gluteraldehyde. Hybridization was performed in a humidified environment overnight at 65˚C with 1μg/ml labeled probe in hybridization buffer (50% formamide, 5X SSC pH4.5, 1% SDS, 50μg/ml yeast tRNA, 50μg/ml heparin). Posthybridization washes were completed at 65˚C using solution I (50% formamide, 5X SSC pH4.5, 1% SDS) and solution III (50% formamide, 2X SSC pH4.5, 0.1% Tween-20) and at room temperature using TBST (TBS containing 1% Tween-20). Brain sections were incubated for >1hr in blocking buffer (TBST, 10% heat-inactivated sheep serum). Immunodetection was carried out in blocking buffer at 4˚C overnight using an alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche) at a 1:2000 dilution. Following antibody incubation extensive TBST washes were performed. Sections were equilibrated in NTMT (100mM Tris-HCl pH 9.5, 100mM NaCl, 50mM $MgCl_2$, 1% Tween-20) prior to colourimetric detection using 2μl/ml NBT/BCIP (Roche) in NTMT. Sections were mounted on Superfrost glass slides (VWR international) and analysed with an Olympus IX51 microscope.

Database pipeline

The database pipeline was used as described in [23]. In summary, the Ensembl database, the International Protein Index and the Mammalian Gene Collection were used to build proteomes for human, mouse, worm and fly. Through automated and manual curation duplicates were removed and only the longest isoforms were kept. This resulted in non-redundant data sets for mouse, human, worm, and fly containing 85991, 74866, 22698, and 16857 sequences, respectively. From an all-against-all Blast search the top 200 hits were used as input for Markov clustering with the MCL program. The output was combined with rich annotation, including predictions of motifs and GPI links, as well as transmembrane and signaling sequences.

Data acquisition and preprocessing

The Allen Brain Atlas (ABA) provides a survey of gene expression in the brain from mid-gestation through aging in different species. Of interest to us is the data on the mouse brain. The detailed process of mouse brain data acquisition is explained in Lein, Hawrylycz [120]. Briefly, ISH data was recorded from coronal or sagittal sections and quantified per unit from a given spatial location for each gene (expression values per voxels). Voxels were assigned regional labels through age-matched anatomic reference atlases. The developmental dataset on the ABA portal (www.brain-map.org) comprises expression values of 2002 genes obtained from ISH images in the sagittal plane across four embryonic (E11.5, E13.5, E15.5 and E18.5) and three early postnatal ages (P4, P14 and P28) [117]. These time-points are developmentally relevant to regional specification/patterning (E11.5), axon pathfinding (E13.5, E15.5 and E18.5),

synaptogenesis (P4 and P14) and cortical plasticity (P28). One animal was used per time-point per gene. Per-region data was quantified from ISH images by combining all voxels with the same regional label. Liscovitch and Chechik [121] extracted data from 36 anatomically delineated regions of the developing brain, which encompass the entire brain. The data is readily available for download at http://chechiklab.biu.ac.il/~lior/cerebellum.html. The data is expressed as expression density. For each brain region R, the expression density is defined as the sum of expressing pixels in R divided by the total number of pixels that intersect R (taken from the Technical White Paper: Informatics Data Processing for the Allen Developing Brain Atlas found at http://developingmouse.brain-map.org/docs/InformaticsDataProcessing.pdf). Each set of values is dependent on specific gene expression. We normalized the data to reflect variations of expression for each gene. A value of 1 indicates average expression levels whereas a value below or above 1 indicates decreased or increased expression levels compared to average, respectively. With this normalised data, we generated the developmental matrix of 1996 genes x 7 time-points as 6 genes had no expression values throughout. These genes are *Chrna1*, *Hoxd9*, *Lef1*, *Nrp2*, *Pbx1* and *Rnd2*.

The genes of the developmental dataset fall into one of these five categories (taken from the Technical White Paper: Allen Developing Mouse Brain Atlas found at http://developingmouse.brain-map.org/docs/Overview.pdf):

1. Transcription factors. Approximately one third of the genes are transcription factors, with extensive coverage of homeobox, basic helix-loop-helix, forkhead, nuclear receptor, high mobility group and POU domain genes.
2. Neuropeptides, neurotransmitters, and their receptors. Extensive coverage of genes in dopaminergic, serotonergic, glutamatergic, and gabaergic signaling, as well as neuropeptides and their receptors.
3. Neuroanatomical marker genes. Characterising region- or cell type specific marker genes over development can provide information about the origins of a brain region or cell type, and may help to identify precursor regions at earlier time-points.
4. Gene ontologies/signaling pathways relevant to brain development. Gene ontologies include axon guidance, receptor tyrosine kinases and their ligands. Pathways include Wnt signaling and Notch signaling pathways.
5. Genes of general interest. This category includes highly studied genes such as common drug targets, ion channels, cell adhesion, genes involved in neurotransmission, G-protein-coupled receptors, or involved in neurodevelopmental diseases, which are expressed in the brain in the adult and/or in development.

We screened protein products for cellular localisation with an emphasis on the extracellular compartment. More specifically, we looked for signal peptides, transmembrane domains and glycophosphatidylinositol (GPI)-anchors. Since genes were listed with their Allen gene name in the dataset, we looked for the MGI symbol and Ensembl gene identifiers (through http://www.ncbi.nlm.nih.gov/gene) to subsequently find their protein counterparts. Where multiple isoforms exist, the longest sequence was used for further analyses. We screened Fasta format sequences through PSORTII to find cellular localisation of the protein products and the number of predicted helices of transmembrane proteins (http://psort.hgc.jp) [122, 123]. PSORTII predicted protein localisation in the following compartments: cytoplasm, cytoskeleton, endoplasmic reticulum, Golgi, mitochondria, nucleus, plasma membrane, peroxisomes, vacuoles and vesicles of the secretory system. In addition, we ran the SignalP prediction program to determine whether a protein contains a signal peptide or not—indicative of extracellular

localisation (http://www.cbs.dtu.dk/services/SignalP/, version 4.1)[124, 125]. A GPI modification site prediction, the big-PI predictor was run to identify protein products with potential GPI-anchors (http://mendel.imp.ac.at/sat/gpi/gpi_server.html, version last modified June 17th 2005) [126–129]. When any of the three programs suggested extracellular localisation, the protein was deemed found in the extracellular compartment. The TMHMM program was run to assess the presence and number of transmembrane domains (http://www.cbs.dtu.dk/services/TMHMM/, version 2.0)[130, 131]. When the number of transmembrane domains varied between PSORTII and TMHMM results, manual curation was performed (unless both programs picked up over two transmembrane domains and the protein was simply considered multi-pass). PFAM (http://pfam.xfam.org)[132] and SMART (http://smart.embl-heidelberg.de, version 2.12.1 released on August 6th 2012)[133, 134], motif recognition programs, allowed screening for important protein domains, which can be amalgamated through InterPro (http://www.ebi.ac.uk/interpro/search/sequence-search, version 49.0 released on November 20th 2014)[135]. For identification of important protein domains we retrieved PFAM, SMART, Interpro and SignalP information through BioMart version 0.7 [136].

Finally, we updated the gene annotations of the dataset from Thompson, Ng (117). Thompson and colleagues had 1385 genes out of the 2002 falling into one or more of these categories: axon guidance pathway, cell adhesion, receptor tyrosine kinases (RTKs) and their ligands, Notch signaling, Wnt signaling, transcription factor activity, basic helix-loop-helix transcription factor, forkhead transcription factor, homeobox transcription factor, nuclear receptor, POU domain genes, neurotransmitter pathway, ion channel, and G-protein coupled receptors (GPCRs). We screened genes for the following gene ontology (GO) term annotations (http://www.geneontology.org)[137] through MartView:

1. Axon Guidance GO:0007411;
2. Synapse GO:0045202;
3. Patterning:—Regionalisation GO:0003002;
   - Pattern Specification Process GO:0007389;
   - Neural Tube Patterning GO:0021532;
   - Rostrocaudal Neural Tube Patterning GO:0021903;
4. Transcription Factor Activity:
   - Protein Binding Transcription Factor Activity GO:0000988;
   - Nucleic Acid Binding Transcription Factor Activity GO:0001071;
   - Sequence-Specific DNA Binding Transcription Factor Activity GO:0003700.

Since many genes had more than one annotation, we grouped them as follows:

- Group 1: axon guidance pathway and cell adhesion;
- Group 2: synapse;
- Group 3: receptor tyrosine kinases and their ligands, and patterning;
- Group 4: neurotransmission pathway (GPCRs, ion channels, gap junctions);
- Group 5: chromatin and transcription factor activity
- Group 6: other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated.

We only had 206 unannotated genes. Each gene belonged to only one group. When a gene fell into more than one group, we prioritised the groups as presented.

Clustering analyses

We performed *k*-means clustering with the Euclidean distance method on the developmental dataset using the clustering module integrated in ArrayPipe (http://www.pathogenomics.ca/arraypipe—version 1.7)[138]. Each clustering analysis was iterated a thousand times.

Enrichment analyses

We performed chi square statistics to determine whether our developmental dataset was normally distributed in regards to cellular localisation and groups across the 10 clusters obtained during *k*-means clustering analyses or enriched in particular clusters. Statistical significance was determined at p-value < 0.05. Where multiple tests were performed on the same dataset, the Bonferroni correction was taken into account.

Supporting information

S1 Table. Conserved candidate connectivity labels. Genes are organised in clusters defined by TRIBE-MCL, as described [23]. Genes from Drosophila melanogaster are in blue, from Caenorhabditis elegans in green and from mouse in black. Signal peptides were predicted using SignalP, transmembrane domains by a consensus between TMHMM and HMMTOP and protein motifs by SMART and PFAM, also as described [23].
(XLS)

S2 Table. Developmental dataset. Absolute values of expression densities for 2002 genes from the devABA database per gene per time-point are shown. NaN stands for Not a Number. Blank cells indicate absence of data in the devABA.
(XLSX)

S3 Table. Developmental dataset and *k*-means clustering. Normalised expression densities per gene per time-point (columns B-H). Thalamic expression densities were obtained for 1996 genes. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; blue, 5. Normalised gene expression was clustered by similar patterns of expression using *k*-means Euclidean distance 6–18, with 1000 iterations (columns I-U). Clusters are sorted at $k = 10$, followed by $k = 11$, followed by $k = 12$ and so on until $k = 18$. Three colour scale of cluster number: red < yellow < green, for better visualisation.
(XLSX)

S4 Table. Developmental dataset protein characterisation. Each gene was matched to its MGI symbol and Ensembl gene identifier. From these, protein sequences were obtained—where multiple isoforms were found, the longest sequence was used for further analyses. Fasta format sequences were screened through PSORT II for cellular localisation (PSORT II prediction and probability: gol, Golgi; cyt, cytoplasmic; csk, cytoskeletal; end, endoplasmic reticulum; ext, extracellular, including cell wall; mit, mitochondrial; nuc, nuclear; per, peroxisomal; pla, plasma membrane; ves, vesicles of secretory system). PSORT II uses two-fold *k-nearest* neighbor (k-NN) algorithm with $k1 = 9$ and $k2 = 23$ [122]. Screens for number of predicted transmembrane domains (TMs) with PSORT II were also performed. SignalP detected signal peptides (sigp) on proteins. Big-PI predictor tested different potential GPI-modification sites and returned the GPI best score, GPI profile score, GPI profile independent score and GPI quality [126–129]. GPI quality of P or S indicated a potential GPI-modification site (P:

predicted; S: second predicted site, more than one predicted site). TMHMM screened for the presence (Transmembrane Domain) and number of transmembrane domains (PredHel) [130, 131]. Columns AE, AF and AG show SMART ID, PFAM ID and InterPro short description, respectively. Genes were screened for additional GO terms (see methods).
(XLSX)

S5 Table. Developmental dataset protein localisation and functional annotations. Normalised expression densities per gene per time-point (Columns B-H). Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Data is sorted by cluster number at $k = 10$ followed by gene name (Column I). Each gene was matched to its MGI symbol (Column J). Protein localisation was assessed after combining the data from the different screens performed as either in the Golgi (gol), cytoplasmic (cyt), cytoskeletal (csk), in the endoplasmic reticulum (end), extracellular including cell wall (ext), mitochondrial (mit), nuclear (nuc), peroxisomal (per), on the plasma membrane (pla) or in vesicles of secretory system (ves; Column K). Moreover, when the protein was deemed in the extracellular compartment, it was found to be either GPI-achored (anchored), secreted, single-pass transmembrane (single-pass) or multi-pass transmembrane (multi-pass; Column L). Functional assessment was grouped in six different categories: 1) axon guidance pathway and cell adhesion; 2) synapse; 3) receptor tyrosine kinases and their ligands, and patterning; 4) neurotransmission pathway (GPCRs, ion channels, gap junctions); 5) chromatin and transcription factor activity; and 6) other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated (Column M). Three colour scale of cluster number and group number: red < yellow < green, for better visualisation (Columns I and M). See text for details.
(XLSX)

S6 Table. Known and potential connectivity genes. Genes are listed by cluster and grouped into those with known roles in thalamocortical connectivity, known functions in axon guidance of synaptogenesis more generally or no such known functions (Other).
(XLSX)

S1 Fig. *In situ* hybridization patterns for immunoglobulin superfamily genes across entire brain at P0. Two coronal sections are shown for *Kirrel3*, *Igsf9b* and *Sdk2*, one rostral and one more caudal. Scale bar: 1 mm.
(TIFF)

S2 Fig. *In situ* hybridization patterns for Odz family genes across entire brain at P0. Two coronal sections are shown for *Odz1*, *Odz2*, *Odz3* and *Odz4*, one rostral and one more caudal. In addition to restricted expression in dorsal thalamus, there is also graded expression of *Odz* genes across cortex (cx) and striatum (str) in differing patterns. Scale bar: 1 mm.
(TIFF)

S3 Fig. *In situ* hybridization patterns for Odz family genes in dorsal thalamus at E15.5. Two coronal sections are shown for *Odz1*, *Odz2*, *Odz3* and *Odz4*, one rostral and one more caudal. Differential expression across the dorsal thalamus is already evident at this stage. The corresponding entire brain sections are shown in Figure S4. Scale bar: 500 μm.
(TIFF)

S4 Fig. *In situ* hybridization patterns for Odz family genes across entire brain at E15.5. Two coronal sections are shown for *Odz1*, *Odz2*, *Odz3* and *Odz4*, one rostral and one more caudal. Scale bar: 1 mm.
(TIFF)

S5 Fig. *In situ* hybridization patterns for Neto family genes across entire brain at P0. Three coronal sections are shown for *Neto1* and *Neto2*, one rostral, one at an intermediate level (mid) and one more caudal. Scale bar: 1 mm.
(TIFF)

S6 Fig. *In situ* hybridization patterns for Neto family genes across entire brain at E15.5. Two coronal sections are shown for *Neto1* and *Neto2*, one rostral and one more caudal. Differential expression across the dorsal thalamus is already evident at this stage. Scale bar: 1 mm.
(TIFF)

S7 Fig. Summaries of expression profiles across *k*-means values. Normalised expression densities were averaged per cluster to see the trends of expression for the results of all clustering analyses from $k = 6$ to 18 (k indicated at top left corner of each table). Clusters were organised chronologically with early peaks of expression at the top and later peaks of expression at the bottom. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. $k = 10$ was used for further analyses; the corresponding tableframe is in bold.
(TIF)

S8 Fig. Image extraction of thalamic gene expression from the devABA. *In situ* hybridization data from a sagittal section of the thalamus at E18.5 obtained from the devABA (A and E). Corresponding section from the anatomic reference atlas (B and F, respectively). Higher magnifications of the thalamus from squared regions in A, B, E and F, respectively (C, D, G and H). Scale bar in A, B, E and F, 880 μm; C, D, G and H, 214 μm. PT (pretectum) and prethalamus (PTh) are labeled within their region bordering the thalamus. p1, prosomere 1 (pretectum and pretectal tegmentum); p1A, alar plate of prosomere 1; p1B, basal plate of prosomere 1; p2, prosomere 2 (thalamus and thalamic tegmentum); p2A, alar plate of prosomere 2; p2B, basal plate of prosomere 2; p3, prosomere 3 (prethalamus and prethalamic tegmentum); p3A, alar plate of prosomere 3; p3B, basal plate of prosomere 3. Note that voxels were assigned regional labels for thalamus and thalamic tegmentum separately.
(TIFF)

S9 Fig. *Odz3 in situ* hybridization compared to devABA data. Two sagittal sections are shown, one lateral and one more medial. Our *in situ* hybridizations are on sections from P0 nenonates, while the devABA sections are from E18.5 embryos. Despite this difference of about a day, there is strikingly good correspondence in expression patterns across the dorsal thalamus, striatum (str), cortex (cx) and other brain regions. Scale bar: 1 mm.
(TIFF)

Acknowledgments

The authors would like to thank Dr. Jackie Dolan for help in managing animal colonies and all the animal facility staff at Trinity College Dublin. This work was supported by grants to KJM from the Wellcome Trust (075264/A/04/Z) and Science Foundation Ireland (07/IN.1/B969 and 09/IN.1/B2614) and to OBB from the Fonds de recherche Santé Québec (23980).

Author Contributions

Conceptualization: OBB TO KJM GT.

Data curation: OBB KH.

Formal analysis: OBB TO KH KJM.

Funding acquisition: OBB KJM GT.

Investigation: OBB TO KH KJM.

Methodology: OBB TO KH KJM.

Project administration: KJM GT.

Software: OBB KH.

Supervision: KJM.

Validation: OBB TO KH KJM.

Visualization: OBB TO KH KJM.

Writing – original draft: OBB TO KH KJM.

Writing – review & editing: OBB TO KH KJM.

References

1. Sherman SM. The thalamus is more than just a relay. Current opinion in neurobiology. 2007; 17 (4):417–22. PubMed Central PMCID: PMC2753250. https://doi.org/10.1016/j.conb.2007.07.003 PMID: 17707635
2. Smith Y, Wichmann T. The cortico-pallidal projection: an additional route for cortical regulation of the basal ganglia circuitry. Movement disorders: official journal of the Movement Disorder Society. 2015; 30(3):293–5. PubMed Central PMCID: PMC4357539.
3. McHaffie JG, Stanford TR, Stein BE, Coizet V, Redgrave P. Subcortical loops through the basal ganglia. Trends in neurosciences. 2005; 28(8):401–7. https://doi.org/10.1016/j.tins.2005.06.006 PMID: 15982753
4. Bostan AC, Dum RP, Strick PL. Cerebellar networks with the cerebral cortex and basal ganglia. Trends in cognitive sciences. 2013; 17(5):241–54. PubMed Central PMCID: PMC3645327. https://doi.org/10.1016/j.tics.2013.03.003 PMID: 23579055
5. Lam YW, Sherman SM. Functional organization of the thalamic input to the thalamic reticular nucleus. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2011; 31(18):6791–9. PubMed Central PMCID: PMC3565464.
6. Clasca F, Rubio-Garrido P, Jabaudon D. Unveiling the diversity of thalamocortical neuron subtypes. The European journal of neuroscience. 2012; 35(10):1524–32. https://doi.org/10.1111/j.1460-9568.2012.08033.x PMID: 22606998
7. Rubio-Garrido P, Perez-de-Manzo F, Porrero C, Galazo MJ, Clasca F. Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. Cerebral cortex. 2009; 19 (10):2380–95. https://doi.org/10.1093/cercor/bhn259 PMID: 19188274
8. Garel S, Lopez-Bendito G. Inputs from the thalamocortical system on axon pathfinding mechanisms. Current opinion in neurobiology. 2014; 27:143–50. https://doi.org/10.1016/j.conb.2014.03.013 PMID: 24742382
9. Molnar Z, Garel S, Lopez-Bendito G, Maness P, Price DJ. Mechanisms controlling the guidance of thalamocortical axons through the embryonic forebrain. The European journal of neuroscience. 2012; 35(10):1573–85. PubMed Central PMCID: PMC4370206. https://doi.org/10.1111/j.1460-9568.2012.08119.x PMID: 22607003
10. Osterhout JA, Josten N, Yamada J, Pan F, Wu SW, Nguyen PL, et al. Cadherin-6 mediates axon-target matching in a non-image-forming visual circuit. Neuron. 2011; 71(4):632–9. PubMed Central PMCID: PMC3513360. https://doi.org/10.1016/j.neuron.2011.07.006 PMID: 21867880
11. Su J, Klemm MA, Josephson AM, Fox MA. Contributions of VLDLR and LRP8 in the establishment of retinogeniculate projections. Neural development. 2013; 8:11. PubMed Central PMCID: PMC3685595. https://doi.org/10.1186/1749-8104-8-11 PMID: 23758727
12. Leamey CA, Merlin S, Lattouf P, Sawatari A, Zhou X, Demel N, et al. Ten_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. PLoS biology. 2007; 5 (9):e241. PubMed Central PMCID: PMC1964777. https://doi.org/10.1371/journal.pbio.0050241 PMID: 17803360

13. Tran H, Sawatari A, Leamey CA. The glycoprotein Ten-m3 mediates topography and patterning of thalamostriatal projections from the parafascicular nucleus in mice. The European journal of neuroscience. 2015; 41(1):55–68. https://doi.org/10.1111/ejn.12767 PMID: 25406022

14. Nakagawa Y, Shimogori T. Diversity of thalamic progenitor cells and postmitotic neurons. The European journal of neuroscience. 2012; 35(10):1554–62. https://doi.org/10.1111/j.1460-9568.2012.08089.x PMID: 22607001

15. Suzuki-Hirano A, Ogawa M, Kataoka A, Yoshida AC, Itoh D, Ueno M, et al. Dynamic spatiotemporal gene expression in embryonic mouse thalamus. The Journal of comparative neurology. 2011; 519 (3):528–43. https://doi.org/10.1002/cne.22531 PMID: 21192082

16. Price DJ, Clegg J, Duocastella XO, Willshaw D, Pratt T. The importance of combinatorial gene expression in early Mammalian thalamic patterning and thalamocortical axonal guidance. Frontiers in neuroscience. 2012; 6:37. PubMed Central PMCID: PMC3304307. https://doi.org/10.3389/fnins.2012.00037 PMID: 22435047

17. Jones EG, Rubenstein JL. Expression of regulatory genes during differentiation of thalamic nuclei in mouse and monkey. The Journal of comparative neurology. 2004; 477(1):55–80. https://doi.org/10.1002/cne.20234 PMID: 15281080

18. Nakagawa Y, O'Leary DD. Combinatorial expression patterns of LIM-homeodomain and other regulatory genes parcellate developing thalamus. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2001; 21(8):2711–25.

19. Lehigh KM, Leonard CE, Baranoski J, Donoghue MJ. Parcellation of the thalamus into distinct nuclei reflects EphA expression and function. Gene expression patterns: GEP. 2013; 13(8):454–63 PubMed Central PMCID: PMC3839050. https://doi.org/10.1016/j.gep.2013.08.002 PMID: 24036135

20. Yuge K, Kataoka A, Yoshida AC, Itoh D, Aggarwal M, Mori S, et al. Region-specific gene expression in early postnatal mouse thalamus. The Journal of comparative neurology. 2011; 519(3):544–61. https://doi.org/10.1002/cne.22532 PMID: 21192083

21. Marcos-Mondejar P, Peregrin S, Li JY, Carlsson L, Tole S, Lopez-Bendito G. The lhx2 transcription factor controls thalamocortical axonal guidance by specific regulation of robo1 and robo2 receptors. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2012; 32(13):4372–85.

22. Chatterjee M, Li JY. Patterning and compartment formation in the diencephalon. Frontiers in neuroscience. 2012; 6:66. PubMed Central PMCID: PMC3349951. https://doi.org/10.3389/fnins.2012.00066 PMID: 22593732

23. Dolan J, Walshe K, Alsbury S, Hokamp K, O'Keeffe S, Okafuji T, et al. The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. BMC genomics. 2007; 8:320. PubMed Central PMCID: PMC2235866. https://doi.org/10.1186/1471-2164-8-320 PMID: 17868438

24. Hansen M, Walmod PS. IGSF9 family proteins. Neurochemical research. 2013; 38(6):1236–51. https://doi.org/10.1007/s11064-013-0999-y PMID: 23417431

25. Woo J, Kwon SK, Nam J, Choi S, Takahashi H, Krueger D, et al. The adhesion protein IgSF9b is coupled to neuroligin 2 via S-SCAM to promote inhibitory synapse development. The Journal of cell biology. 2013; 201(6):929–44. PubMed Central PMCID: PMC3678166. https://doi.org/10.1083/jcb.201209132 PMID: 23751499

26. Mishra A, Traut MH, Becker L, Klopstock T, Stein V, Klein R. Genetic evidence for the adhesion protein IgSF9/Dasm1 to regulate inhibitory synapse development independent of its intracellular domain. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2014; 34(12):4187–99.

27. Fischbach KF, Linneweber GA, Andlauer TF, Hertenstein A, Bonengel B, Chaudhary K. The irre cell recognition module (IRM) proteins. Journal of neurogenetics. 2009; 23(1–2):48–67. https://doi.org/10.1080/01677060802471668 PMID: 19132596

28. Shen K, Fetter RD, Bargmann CI. Synaptic specificity is generated by the synaptic guidepost protein SYG-2 and its receptor, SYG-1. Cell. 2004; 116(6):869–81. PMID: 15035988

29. Prince JE, Brignall AC, Cutforth T, Shen K, Cloutier JF. Kirrel3 is required for the coalescence of vomeronasal sensory neuron axons into glomeruli and for male-male aggression. Development. 2013; 140 (11):2398–408. PubMed Central PMCID: PMC3653560. https://doi.org/10.1242/dev.087262 PMID: 23637329

30. Martin EA, Muralidhar S, Wang Z, Cervantes DC, Basu R, Taylor MR, et al. The intellectual disability gene Kirrel3 regulates target-specific mossy fiber synapse development in the hippocampus. eLife. 2015; 4:e09395. PubMed Central PMCID: PMC4642954. https://doi.org/10.7554/eLife.09395 PMID: 26575286

31. Nguyen DN, Liu Y, Litsky ML, Reinke R. The sidekick gene, a member of the immunoglobulin superfamily, is required for pattern formation in the Drosophila eye. Development. 1997; 124(17):3303–12. PMID: 9310325

32. Krishnaswamy A, Yamagata M, Duan X, Hong YK, Sanes JR. Sidekick 2 directs formation of a retinal circuit that detects differential motion. Nature. 2015; 524(7566):466–70. PubMed Central PMCID: PMC4552609. https://doi.org/10.1038/nature14682 PMID: 26287463

33. Yamagata M, Sanes JR. Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. Nature. 2008; 451(7177):465–9. https://doi.org/10.1038/nature06469 PMID: 18216854

34. Mosca TJ. On the Teneurin track: a new synaptic organization molecule emerges. Frontiers in cellular neuroscience. 2015; 9:204. PubMed Central PMCID: PMC4444827. https://doi.org/10.3389/fncel.2015.00204 PMID: 26074772

35. Leamey CA, Sawatari A. The teneurins: new players in the generation of visual topography. Seminars in cell & developmental biology. 2014; 35:173–9.

36. Copits BA, Swanson GT. Dancing partners at the synapse: auxiliary subunits that shape kainate receptor function. Nature reviews Neuroscience. 2012; 13(10):675–86. PubMed Central PMCID: PMC3520510. https://doi.org/10.1038/nrn3335 PMID: 22948074

37. Kim YJ, Bao H, Bonanno L, Zhang B, Serpe M. Drosophila Neto is essential for clustering glutamate receptors at the neuromuscular junction. Genes & development. 2012; 26(9):974–87. PubMed Central PMCID: PMC3347794.

38. Hintsch G, Zurlinden A, Meskenaite V, Steuble M, Fink-Widmer K, Kinter J, et al. The calsyntenins—a family of postsynaptic membrane proteins with distinct neuronal expression patterns. Molecular and cellular neurosciences. 2002; 21(3):393–409. PMID: 12498782

39. Um JW, Pramanik G, Ko JS, Song MY, Lee D, Kim H, et al. Calsyntenins function as synaptogenic adhesion molecules in concert with neurexins. Cell reports. 2014; 6(6):1096–109. PubMed Central PMCID: PMC4101519. https://doi.org/10.1016/j.celrep.2014.02.010 PMID: 24613359

40. Pfeiffenberger C, Yamada J, Feldheim DA. Ephrin-As and patterned retinal activity act together in the development of topographic maps in the primary visual system. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2006; 26(50):12873–84. PubMed Central PMCID: PMC3664553.

41. Dufour A, Seibt J, Passante L, Depaepe V, Ciossek T, Frisen J, et al. Area specificity and topography of thalamocortical projections are controlled by ephrin/Eph genes. Neuron. 2003; 39(3):453–65. PMID: 12895420

42. Uziel D, Garcez P, Lent R, Peuckert C, Niehage R, Weth F, et al. Connecting thalamus and cortex: the role of ephrins. The anatomical record Part A, Discoveries in molecular, cellular, and evolutionary biology. 2006; 288(2):135–42. https://doi.org/10.1002/ar.a.20286 PMID: 16411249

43. Tai AX, Kromer LF. Corticofugal projections from medial primary somatosensory cortex avoid EphA7-expressing neurons in striatum and thalamus. Neuroscience. 2014; 274:409–18. https://doi.org/10.1016/j.neuroscience.2014.05.039 PMID: 24909897

44. Torii M, Levitt P. Dissociation of corticothalamic and thalamocortical axon targeting by an EphA7-mediated mechanism. Neuron. 2005; 48(4):563–75. https://doi.org/10.1016/j.neuron.2005.09.021 PMID: 16301174

45. Gerstmann K, Pensold D, Symmank J, Khundadze M, Hubner CA, Bolz J, et al. Thalamic afferents influence cortical progenitors via ephrin A5-EphA4 interactions. Development. 2015; 142(1):140–50. https://doi.org/10.1242/dev.104927 PMID: 25480914

46. Robichaux MA, Chenaux G, Ho HY, Soskis MJ, Dravis C, Kwan KY, et al. EphB receptor forward signaling regulates area-specific reciprocal thalamic and cortical axon pathfinding. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(6):2188–93 PubMed Central PMCID: PMC3926086. https://doi.org/10.1073/pnas.1324215111 PMID: 24453220

47. Yun ME, Johnson RR, Antic A, Donoghue MJ. EphA family gene expression in the developing mouse neocortex: regional patterns reveal intrinsic programs and extrinsic influence. The Journal of comparative neurology. 2003; 456(3):203–16. https://doi.org/10.1002/cne.10498 PMID: 12528186

48. Suzuki SC, Inoue T, Kimura Y, Tanaka T, Takeichi M. Neuronal circuits are subdivided by differential expression of type-II classic cadherins in postnatal mouse brains. Molecular and cellular neurosciences. 1997; 9(5–6):433–47. https://doi.org/10.1006/mcne.1997.0626 PMID: 9361280

49. Poskanzer K, Needleman LA, Bozdagi O, Huntley GW. N-cadherin regulates ingrowth and laminar targeting of thalamocortical axons. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2003; 23(6):2294–305. PubMed Central PMCID: PMC4415263.

50. Gil OD, Needleman L, Huntley GW. Developmental patterns of cadherin expression and localization in relation to compartmentalized thalamocortical terminations in rat barrel cortex. The Journal of comparative neurology. 2002; 453(4):372–88. https://doi.org/10.1002/cne.10424 PMID: 12389209

51. Kim SY, Chung HS, Sun W, Kim H. Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. Neuroscience. 2007; 147(4):996–1021. https://doi.org/10.1016/j.neuroscience.2007.03.052 PMID: 17614211

52. Uemura M, Nakao S, Suzuki ST, Takeichi M, Hirano S. OL-Protocadherin is essential for growth of striatal axons and thalamocortical projections. Nature neuroscience. 2007; 10(9):1151–9. https://doi.org/10.1038/nn1960 PMID: 17721516

53. Carcea I, Patil SB, Robison AJ, Mesias R, Huntsman MM, Froemke RC, et al. Maturation of cortical circuits requires Semaphorin 7A. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(38):13978–83. PubMed Central PMCID: PMC4183324. https://doi.org/10.1073/pnas.1408680111 PMID: 25201975

54. Fukunishi A, Maruyama T, Zhao H, Tiwari M, Kang S, Kumanogoh A, et al. The action of Semaphorin7A on thalamocortical axon branching. Journal of neurochemistry. 2011; 118(6):1008–15. https://doi.org/10.1111/j.1471-4159.2011.07390.x PMID: 21781117

55. Leighton PA, Mitchell KJ, Goodrich LV, Lu X, Pinson K, Scherz P, et al. Defining brain wiring patterns and mechanisms through gene trapping in mice. Nature. 2001; 410(6825):174–9. https://doi.org/10.1038/35065539 PMID: 11242070

56. Little GE, Lopez-Bendito G, Runker AE, Garcia N, Pinon MC, Chedotal A, et al. Specificity and plasticity of thalamocortical connections in Sema6A mutant mice. PLoS biology. 2009; 7(4):e98. PubMed Central PMCID: PMC2672616. https://doi.org/10.1371/journal.pbio.1000098 PMID: 19402755

57. Demyanenko GP, Riday TT, Tran TS, Dalal J, Darnell EP, Brennaman LH, et al. NrCAM deletion causes topographic mistargeting of thalamocortical axons to the visual cortex and disrupts visual acuity. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2011; 31(4):1545–58. PubMed Central PMCID: PMC3037548.

58. Wright AG, Demyanenko GP, Powell A, Schachner M, Enriquez-Barreto L, Tran TS, et al. Close homolog of L1 and neuropilin 1 mediate guidance of thalamocortical axons at the ventral telencephalon. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2007; 27(50):13667–79.

59. Guijarro P, Wang Y, Ying Y, Yao Y, Jieyi X, Yuan X. In vivo knockdown of cKit impairs neuronal migration and axonal extension in the cerebral cortex. Developmental neurobiology. 2013; 73(12):871–87. https://doi.org/10.1002/dneu.22107 PMID: 23843227

60. Widenfalk J, Nosrat C, Tomac A, Westphal H, Hoffer B, Olson L. Neurturin and glial cell line-derived neurotrophic factor receptor-beta (GDNFR-beta), novel proteins related to GDNF and GDNFR-alpha with specific cellular patterns of expression suggesting roles in the developing and adult nervous system and in peripheral organs. The Journal of neuroscience: the official journal of the Society for Neuroscience. 1997; 17(21):8506–19.

61. Lush ME, Ma L, Parada LF. TrkB signaling regulates the developmental maturation of the somatosensory cortex. International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience. 2005; 23(6):523–36.

62. Cheng J, Wu K, Armanini M, O'Rourke N, Dowbenko D, Lasky LA. A novel protein-tyrosine phosphatase related to the homotypically adhering kappa and mu receptors. The Journal of biological chemistry. 1997; 272(11):7264–77. PMID: 9054423

63. Taniguchi Y, London R, Schinkmann K, Jiang S, Avraham H. The receptor protein tyrosine phosphatase, PTP-RO, is upregulated during megakaryocyte differentiation and Is associated with the c-Kit receptor. Blood. 1999; 94(2):539–49. PMID: 10397721

64. Johnson KG, Van Vactor D. Receptor protein tyrosine phosphatases in nervous system development. Physiological reviews. 2003; 83(1):1–24. https://doi.org/10.1152/physrev.00016.2002 PMID: 12506125

65. Weiner JA, Koo SJ, Nicolas S, Fraboulet S, Pfaff SL, Pourquie O, et al. Axon fasciculation defects and retinal dysplasias in mice lacking the immunoglobulin superfamily adhesion molecule BEN/ALCAM/SC1. Molecular and cellular neurosciences. 2004; 27(1):59–69. https://doi.org/10.1016/j.mcn.2004.06.005 PMID: 15345243

66. Buhusi M, Demyanenko GP, Jannie KM, Dalal J, Darnell EP, Weiner JA, et al. ALCAM regulates mediolateral retinotopic mapping in the superior colliculus. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2009; 29(50):15630–41.

67. Fogel AI, Akins MR, Krupp AJ, Stagi M, Stein V, Biederer T. SynCAMs organize synapses through heterophilic adhesion. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2007; 27(46):12516–30.

68. Frei JA, Stoeckli ET. SynCAMs extend their functions beyond the synapse. The European journal of neuroscience. 2014; 39(11):1752–60. https://doi.org/10.1111/ejn.12544 PMID: 24628990

69. Toth AB, Terauchi A, Zhang LY, Johnson-Venkatesh EM, Larsen DJ, Sutton MA, et al. Synapse maturation by activity-dependent ectodomain shedding of SIRPalpha. Nature neuroscience. 2013; 16 (10):1417–25. PubMed Central PMCID: PMC3820962. https://doi.org/10.1038/nn.3516 PMID: 24036914

70. Hsieh CP, Chang WT, Lee YC, Huang AM. Deficits in cerebellar granule cell development and social interactions in CD47 knockout mice. Developmental neurobiology. 2015; 75(5):463–84. https://doi.org/10.1002/dneu.22236 PMID: 25288019

71. Shimoda Y, Watanabe K. Contactins: emerging key roles in the development and function of the nervous system. Cell adhesion & migration. 2009; 3(1):64–70. PubMed Central PMCID: PMC2675151.

72. Sakurai K, Toyoshima M, Ueda H, Matsubara K, Takeda Y, Karagogeos D, et al. Contribution of the neural cell recognition molecule NB-3 to synapse formation between parallel fibers and Purkinje cells in mouse. Developmental neurobiology. 2009; 69(12):811–24. https://doi.org/10.1002/dneu.20742 PMID: 19672956

73. Sakurai K, Toyoshima M, Takeda Y, Shimoda Y, Watanabe K. Synaptic formation in subsets of glutamatergic terminals in the mouse hippocampal formation is affected by a deficiency in the neural cell recognition molecule NB-3. Neuroscience letters. 2010; 473(2):102–6. https://doi.org/10.1016/j.neulet.2010.02.027 PMID: 20176085

74. Lee S, Takeda Y, Kawano H, Hosoya H, Nomoto M, Fujimoto D, et al. Expression and regulation of a gene encoding neural recognition molecule NB-3 of the contactin/F3 subgroup in mouse brain. Gene. 2000; 245(2):253–66. PMID: 10717476

75. Takeda Y, Akasaka K, Lee S, Kobayashi S, Kawano H, Murayama S, et al. Impaired motor coordination in mice lacking neural recognition molecule NB-3 of the contactin/F3 subgroup. Journal of neurobiology. 2003; 56(3):252–65. https://doi.org/10.1002/neu.10222 PMID: 12884264

76. Takeuchi A, Hamasaki T, Litwack ED, O'Leary DD. Novel IgCAM, MDGA1, expressed in unique cortical area- and layer-specific patterns and transiently by distinct forebrain populations of Cajal-Retzius neurons. Cerebral cortex. 2007; 17(7):1531–41. https://doi.org/10.1093/cercor/bhl064 PMID: 16959869

77. Ishikawa T, Gotoh N, Murayama C, Abe T, Iwashita M, Matsuzaki F, et al. IgSF molecule MDGA1 is involved in radial migration and positioning of a subset of cortical upper-layer neurons. Developmental dynamics: an official publication of the American Association of Anatomists. 2011; 240(1):96–107.

78. Lee K, Kim Y, Lee SJ, Qiang Y, Lee D, Lee HW, et al. MDGAs interact selectively with neuroligin-2 but not other neuroligins to regulate inhibitory synapse development. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(1):336–41. PubMed Central PMCID: PMC3538197. https://doi.org/10.1073/pnas.1219987110 PMID: 23248271

79. Pettem KL, Yokomaku D, Takahashi H, Ge Y, Craig AM. Interaction between autism-linked MDGAs and neuroligins suppresses inhibitory synapse development. The Journal of cell biology. 2013; 200 (3):321–36. PubMed Central PMCID: PMC3563690. https://doi.org/10.1083/jcb.201206028 PMID: 23358245

80. de Wit J, Hong W, Luo L, Ghosh A. Role of leucine-rich repeat proteins in the development and function of neural circuits. Annual review of cell and developmental biology. 2011; 27:697–729 https://doi.org/10.1146/annurev-cellbio-092910-154111 PMID: 21740233

81. Kegel L, Aunin E, Meijer D, Bermingham JR. LGI proteins in the nervous system. ASN neuro. 2013; 5 (3):167–81. PubMed Central PMCID: PMC3691968. https://doi.org/10.1042/AN20120095 PMID: 23713523

82. Fukata Y, Lovero KL, Iwanaga T, Watanabe A, Yokoi N, Tabuchi K, et al. Disruption of LGI1-linked synaptic complex causes abnormal synaptic transmission and epilepsy. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(8):3799–804. PubMed Central PMCID: PMC2840530. https://doi.org/10.1073/pnas.0914537107 PMID: 20133599

83. Fukata Y, Adesnik H, Iwanaga T, Bredt DS, Nicoll RA, Fukata M. Epilepsy-related ligand/receptor complex LGI1 and ADAM22 regulate synaptic transmission. Science. 2006; 313(5794):1792–5. https://doi.org/10.1126/science.1129947 PMID: 16990550

84. Seppala EH, Jokinen TS, Fukata M, Fukata Y, Webster MT, Karlsson EK, et al. LGI2 truncation causes a remitting focal epilepsy in dogs. PLoS genetics. 2011; 7(7):e1002194. PubMed Central PMCID: PMC3145619. https://doi.org/10.1371/journal.pgen.1002194 PMID: 21829378

85. Andreae LC, Lumsden A, Gilthorpe JD. Chick Lrrn2, a novel downstream effector of Hoxb1 and Shh, functions in the selective targeting of rhombomere 4 motor neurons. Neural development. 2009; 4:27. PubMed Central PMCID: PMC2716342. https://doi.org/10.1186/1749-8104-4-27 PMID: 19602272

86. Tossell K, Andreae LC, Cudmore C, Lang E, Muthukrishnan U, Lumsden A, et al. Lrrn1 is required for formation of the midbrain-hindbrain boundary and organiser through regulation of affinity differences between midbrain and hindbrain cells in chick. Developmental biology. 2011; 352(2):341–52. PubMed Central PMCID: PMC3084456. https://doi.org/10.1016/j.ydbio.2011.02.002 PMID: 21315708

87. Aruga J. Slitrk6 expression profile in the mouse embryo and its relationship to that of Nlrr3. Gene expression patterns: GEP. 2003; 3(6):727–33. PMID: 14643680

88. Dickendesher TL, Baldwin KT, Mironova YA, Koriyama Y, Raiker SJ, Askew KL, et al. NgR1 and NgR3 are receptors for chondroitin sulfate proteoglycans. Nature neuroscience. 2012; 15(5):703–12. PubMed Central PMCID: PMC3337880. https://doi.org/10.1038/nn.3070 PMID: 22406547

89. Wills ZP, Mandel-Brehm C, Mardinly AR, McCord AE, Giger RJ, Greenberg ME. The nogo receptor family restricts synapse number in the developing hippocampus. Neuron. 2012; 73(3):466–81. PubMed Central PMCID: PMC3532882. https://doi.org/10.1016/j.neuron.2011.11.029 PMID: 22325200

90. Uemura T, Lee SJ, Yasumura M, Takeuchi T, Yoshida T, Ra M, et al. Trans-synaptic interaction of GluRdelta2 and Neurexin through Cbln1 mediates synapse formation in the cerebellum. Cell. 2010; 141(6):1068–79. https://doi.org/10.1016/j.cell.2010.04.035 PMID: 20537373

91. Wei P, Pattarini R, Rong Y, Guo H, Bansal PK, Kusnoor SV, et al. The Cbln family of proteins interact with multiple signaling pathways. Journal of neurochemistry. 2012; 121(5):717–29. PubMed Central PMCID: PMC3342465. https://doi.org/10.1111/j.1471-4159.2012.07648.x PMID: 22220752

92. Miura E, Iijima T, Yuzaki M, Watanabe M. Distinct expression of Cbln family mRNAs in developing and adult mouse brains. The European journal of neuroscience. 2006; 24(3):750–60. https://doi.org/10.1111/j.1460-9568.2006.04950.x PMID: 16930405

93. Matsukawa H, Akiyoshi-Nishimura S, Zhang Q, Lujan R, Yamaguchi K, Goto H, et al. Netrin-G/NGL complexes encode functional synaptic diversification. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2014; 34(47):15779–92.

94. Nakashiba T, Nishimura S, Ikeda T, Itohara S. Complementary expression and neurite outgrowth activity of netrin-G subfamily members. Mechanisms of development. 2002; 111(1–2):47–60. PMID: 11804778

95. Karayannis T, Au E, Patel JC, Kruglikov I, Markx S, Delorme R, et al. Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. Nature. 2014; 511(7508):236–40. PubMed Central PMCID: PMC4281262. https://doi.org/10.1038/nature13248 PMID: 24870235

96. de Wit J, O'Sullivan ML, Savas JN, Condomitti G, Caccese MC, Vennekens KM, et al. Unbiased discovery of glypican as a receptor for LRRTM4 in regulating excitatory synapse development. Neuron. 2013; 79(4):696–711. PubMed Central PMCID: PMC4003527. https://doi.org/10.1016/j.neuron.2013.06.049 PMID: 23911103

97. Budnik V, Salinas PC. Wnt signaling during synaptic development and plasticity. Current opinion in neurobiology. 2011; 21(1):151–9. PubMed Central PMCID: PMC3499977. https://doi.org/10.1016/j.conb.2010.12.002 PMID: 21239163

98. Sahores M, Gibb A, Salinas PC. Frizzled-5, a receptor for the synaptic organizer Wnt7a, regulates activity-mediated synaptogenesis. Development. 2010; 137(13):2215–25. PubMed Central PMCID: PMC2882138. https://doi.org/10.1242/dev.046722 PMID: 20530549

99. Liu C, Wang Y, Smallwood PM, Nathans J. An essential role for Frizzled5 in neuronal survival in the parafascicular nucleus of the thalamus. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28(22):5641–53.

100. Wilson PM, Fryer RH, Fang Y, Hatten ME. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30(25):8529–40. PubMed Central PMCID: PMC2905051.

101. Garcia-Gallastegui P, Ibarretxe G, Garcia-Ramirez JJ, Baladron V, Aurrekoetxea M, Nueda ML, et al. DLK1 regulates branching morphogenesis and parasympathetic innervation of salivary glands through inhibition of NOTCH signalling. Biology of the cell / under the auspices of the European Cell Biology Organization. 2014; 106(8):237–53.

102. Fukazawa N, Yokoyama S, Eiraku M, Kengaku M, Maeda N. Receptor type protein tyrosine phosphatase zeta-pleiotrophin signaling controls endocytic trafficking of DNER that regulates neuritogenesis. Molecular and cellular biology. 2008; 28(14):4494–506. PubMed Central PMCID: PMC2447117. https://doi.org/10.1128/MCB.00074-08 PMID: 18474614

103. Dessaud E, Salaun D, Gayet O, Chabbert M, deLapeyriere O. Identification of lynx2, a novel member of the ly-6/neurotoxin superfamily, expressed in neuronal subpopulations during mouse development. Molecular and cellular neurosciences. 2006; 31(2):232–42. https://doi.org/10.1016/j.mcn.2005.09.010 PMID: 16236524

104. Burazin TC, Gundlach AL. Localization of GDNF/neurturin receptor (c-ret, GFRalpha-1 and alpha-2) mRNAs in postnatal rat brain: differential regional and temporal expression in hippocampus, cortex and cerebellum. Brain research Molecular brain research. 1999; 73(1–2):151–71. PMID: 10581409

105. Golden JP, DeMaro JA, Osborne PA, Milbrandt J, Johnson EM Jr. Expression of neurturin, GDNF, and GDNF family-receptor mRNA in the developing and mature mouse. Experimental neurology. 1999; 158(2):504–28. https://doi.org/10.1006/exnr.1999.7127 PMID: 10415156

106. Sato H, Fukutani Y, Yamamoto Y, Tatara E, Takemoto M, Shimamura K, et al. Thalamus-derived molecules promote survival and dendritic growth of developing cortical neurons. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2012; 32(44):15388–402.

107. Kao TJ, Kania A. Ephrin-mediated cis-attenuation of Eph receptor signaling is essential for spinal motor axon guidance. Neuron. 2011; 71(1):76–91. https://doi.org/10.1016/j.neuron.2011.05.031 PMID: 21745639

108. Sun LO, Brady CM, Cahill H, Al-Khindi T, Sakuta H, Dhande OS, et al. Functional assembly of accessory optic system circuitry critical for compensatory eye movements. Neuron. 2015; 86(4):971–84. PubMed Central PMCID: PMC4441577. https://doi.org/10.1016/j.neuron.2015.03.064 PMID: 25959730

109. Haklai-Topper L, Mlechkovich G, Savariego D, Gokhman I, Yaron A. Cis interaction between Semaphorin6A and Plexin-A4 modulates the repulsive response to Sema6A. The EMBO journal. 2010; 29 (15):2635–45. PubMed Central PMCID: PMC2928682. https://doi.org/10.1038/emboj.2010.147 PMID: 20606624

110. Perez-Branguli F, Zagar Y, Shanley DK, Graef IA, Chedotal A, Mitchell KJ. Reverse Signaling by Semaphorin-6A Regulates Cellular Aggregation and Neuronal Morphology. PloS one. 2016; 11(7): e0158686. PubMed Central PMCID: PMC4938514. https://doi.org/10.1371/journal.pone.0158686 PMID: 27392094

111. Sotomayor M, Gaudet R, Corey DP. Sorting out a promiscuous superfamily: towards cadherin connectomics. Trends in cell biology. 2014; 24(9):524–36. PubMed Central PMCID: PMC4294768. https://doi.org/10.1016/j.tcb.2014.03.007 PMID: 24794279

112. Basu R, Taylor MR, Williams ME. The classic cadherins in synaptic specificity. Cell adhesion & migration. 2015; 9(3):193–201. PubMed Central PMCID: PMC4594527.

113. Emond MR, Biswas S, Blevins CJ, Jontes JD. A complex of Protocadherin-19 and N-cadherin mediates a novel mechanism of cell adhesion. The Journal of cell biology. 2011; 195(7):1115–21. PubMed Central PMCID: PMC3246890. https://doi.org/10.1083/jcb.201108115 PMID: 22184198

114. Schmid RS, Maness PF. L1 and NCAM adhesion molecules as signaling coreceptors in neuronal migration and process outgrowth. Current opinion in neurobiology. 2008; 18(3):245–50. PubMed Central PMCID: PMC2633433. https://doi.org/10.1016/j.conb.2008.07.015 PMID: 18760361

115. Yang X, Hou D, Jiang W, Zhang C. Intercellular protein-protein interactions at synapses. Protein & cell. 2014; 5(6):420–44. PubMed Central PMCID: PMC4026422.

116. de Wit J, Ghosh A. Specification of synaptic connectivity by cell surface interactions. Nature reviews Neuroscience. 2016; 17(1):22–35. https://doi.org/10.1038/nrn.2015.3 PMID: 26656254

117. Thompson CL, Ng L, Menon V, Martinez S, Lee CK, Glattfelder K, et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron. 2014; 83(2):309–23. PubMed Central PMCID: PMC4319559. https://doi.org/10.1016/j.neuron.2014.05.033 PMID: 24952961

118. de Wit J, Ghosh A. Control of neural circuit formation by leucine-rich repeat proteins. Trends in neurosciences. 2014; 37(10):539–50. PubMed Central PMCID: PMC4189993. https://doi.org/10.1016/j.tins.2014.07.004 PMID: 25131359

119. Riddle RD, Johnson RL, Laufer E, Tabin C. Sonic hedgehog mediates the polarizing activity of the ZPA. Cell. 1993; 75(7):1401–16. PMID: 8269518

120. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007; 445(7124):168–76. https://doi.org/10.1038/nature05453 PMID: 17151600

121. Liscovitch N, Chechik G. Specialization of gene expression during mouse brain development. PLoS computational biology. 2013; 9(9):e1003185. PubMed Central PMCID: PMC3777910. https://doi.org/10.1371/journal.pcbi.1003185 PMID: 24068900

122. Horton P, Nakai K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. Proceedings / International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology. 1997; 5:147–52.

123. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends in biochemical sciences. 1999; 24(1):34–6. PMID: 10087920

124. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering. 1997; 10(1):1–6. PMID: 9051728

125. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods. 2011; 8(10):785–6. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

126. Eisenhaber B, Bork P, Eisenhaber F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein engineering. 1998; 11(12):1155–61. PMID: 9930665

127. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. Journal of molecular biology. 1999; 292(3):741–58. https://doi.org/10.1006/jmbi.1999.3069 PMID: 10497036

128. Eisenhaber B, Bork P, Yuan Y, Loffler G, Eisenhaber F. Automated annotation of GPI anchor sites: case study C. elegans. Trends in biochemical sciences. 2000; 25(7):340–1. PMID: 10871885

129. Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. Protein engineering. 1999; 12(5):387–94. PMID: 10360979

130. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

131. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998; 6:175–82. PMID: 9783223

132. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic acids research. 2014; 42(Database issue):D222–30. PubMed Central PMCID: PMC3965110. https://doi.org/10.1093/nar/gkt1223 PMID: 24288371

133. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic acids research. 2014.

134. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(11):5857–64. PubMed Central PMCID: PMC34487. PMID: 9600884

135. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: an integrated documentation resource for protein families, domains and functional sites. Briefings in bioinformatics. 2002; 3(3):225–35. PMID: 12230031

136. Kasprzyk A. BioMart: driving a paradigm change in biological data management. Database: the journal of biological databases and curation. 2011; 2011:bar049. PubMed Central PMCID: PMC3215098. https://doi.org/10.1093/database/bar049 PMID: 22083790

137. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25(1):25–9. PubMed Central PMCID: PMC3037419. https://doi.org/10.1038/75556 PMID: 10802651

138. Hokamp K, Roche FM, Acab M, Rousseau ME, Kuo B, Goode D, et al. ArrayPipe: a flexible processing pipeline for microarray data. Nucleic acids research. 2004; 32(Web Server issue):W457–9. PubMed Central PMCID: PMC441584. https://doi.org/10.1093/nar/gkh446 PMID: 15215429